Ethical Approaches to Closed Messaging Research

Considerations in Democratic Contexts

Connie Moon Sehat and Tarunima Prabhakar, with Aleksei Kaminski

March 15, 2021

An initiative of

MISINFOCON
МИСИНФОКОН
Trust, Verification, Fact Checking & Beyond

THE CARTER CENTER

Contents

Executive Summary **1**

Background: The Reach of Closed Messaging Systems **3**

Motivations for Analyzing Political Discussions within Closed Spaces **6**

The Private-Public Conundrum for Researchers **9**

Closed Messaging Spaces are Social *9*
Yet Companies Communicate and Users Expect Privacy *11*

Models of Research Investigation **13**

Baseline Model: Smaller Investigations under Informed Consent 14

Model 1: Voluntary Contributions 15

Model 2: Join Groups through Invitation with a Public/Publicized Purpose 16

Model 3: Join "Public" Groups Without Disclosing Research Intent 18

Model 4: Join 'Public' Groups while Disclosing Research Intent 21

Design Questions and Choices for Researchers **21**

Parameters: Exactly When Is a Closed Message Chat 'Public'? 23

Technological Signals: Publicly Indexed or Available Invitations, Group classification *23*
Public Interest: Topic and Legal Considerations *23*
Group Size and Expectation *25*
Overlapping Methodological Definitions of Public and Private *25*

Parameters: Who Does the Data Belong To? 26

Security: What is the Plan for Storing Closed Messaging Research Data in a Secure Fashion? 27

Security: How Might the Research Project Affect Group Members and Their Conversations? 27

Obligations: What are the Obligations for Researcher Disclosure and Informed Consent? 28

Obligations: When Should Researchers Inform or Report Back to the Groups Involved? 29

Next Steps: Public Discussions amid an Evolving Technology **31**

Appendix A: Closed Messaging Data Collection and Sharing Practices **35**

Signing Up on The App 36

Discovering Relevant Groups 37

After Joining A Group Conversation 39

Disclosure Norms *39*

Deciding to Leave a Group 39
Extracting and Collating Data for Analysis 40
Method of Data Collection 40
De-identification and Anonymization 41
Backup 42
Reproducibility 42
Scenarios 42
Scenario One: Minimizing Data Collection 43
Scenario Two: Journalism Investigation 43
Scenario Three: Quantitative Project 44
Discussion 44
Appendix B: Examples of Considerations, Practices, and Questions Regarding Closed Message Research 45
Considerations from Social Media Data Stewardship, 2016 45
Considerations from Frankel, 2018 46
Considerations from Sehat, 2018 46
Considerations and Practices from Bengani, 2019 47
Questions from Dotto et al., 2019 47
Considerations from Gibson, 2020 48
Considerations from Kaufmann and Peil, 2020 48
Appendix C: References and Additional Resources 50
Additional Resources 59
Acknowledgements 61

Executive Summary

"Closed" messaging apps such as WhatsApp, Signal and Telegram have grown in reach and adoption in recent years and have transformed elections-related communications. However, understanding their impact upon larger public discussion poses a conundrum for researchers. On the one hand, widely accessible conversations of public importance exist in these spaces, but on the other, message encryption challenges existing professional ethics of access and collection. In order to study what messages are flowing through these channels, analysts and researchers must join potentially private chat groups. This raises the question: What considerations should be taken into account when conducting research into closed messaging spaces within democratic contexts, in which the individual right to privacy also prevails?[1]

To better understand the decisions that researchers face, this paper presents research practices taking place within these applications. It focuses on projects with election-related themes that parse the content of messages. Two investigating groups are of interest: *public-interest organizations* focused on human rights, democratic elections, and fact-checking that seek real-time or more immediate impact based upon their findings, and *academics* who want to provide longer-term, systematic analysis of the political dynamics created by these closed conversations. Sometimes these groups work in collaboration with each other as well as with other types of researchers, such as those based at technology companies.[2]

When it comes to studies involving the systematic collection of message data, there are at least four models of research practice. These models are summarized below:

[1] While messenger app usage in countries like China is quite dominant, this paper focuses instead on country contexts with multiple party systems and liberal democratic frameworks that have different requirements for public versus private conversation.

[2] While this topic has relevance for corporate research and development approaches, this paper is focused upon public, outward-facing research results. Two examples of collaboration with corporations are included in the review.

- Baseline model: Smaller Investigations under Informed Consent
- Model 1: Voluntary Contributions (e.g., tiplines)
- Model 2: Joining Groups through Invitation with Public/Publicized Purpose
- Model 3: Joining "Public" Groups Without Disclosing Research Intent
- Model 4: Joining "Public" Groups While Disclosing Research Intent

First, we share background about the reach of closed messaging applications and the motivation for these research efforts. Summaries of the four models follow. Based on the models, the final section of the paper offers a set of key questions based on public parameters, participant safety, and researcher obligations, to help clarify the ethical tensions posed by investigations into closed message discussions:

- Parameters: When exactly is a closed message chat "public"?
- Parameters: Who does the data belong to?
- Security: What is the plan for storing closed messaging data in a secure fashion?
- Security: How might the research project affect group members and their conversations?
- Obligations: What are the obligations for research disclosure and informed consent?
- Obligations: When should researchers inform or report back to the groups involved?

In addition, a data collection guide is included at the end of the paper, which pairs practical steps with key ethical considerations.

Background: The Reach of Closed Messaging Systems

Closed messaging applications or "messenger apps" such as WhatsApp, Facebook Messenger, and WeChat are now among the world's leading social media platforms and applications; according to marketing analytics reports, these applications were among the top 10 most-used social platforms at the onset of 2020.[3] Different from the openness of Twitter or even the default Facebook experience, these "private chat" spaces for personal exchange are having a larger impact upon public discourse than originally envisioned.[4]

There are a few reasons for this broad effect: Closed messaging chats —which range in size from one-on-one discussion to groups of over 200 or even 100,000 accounts — can be much larger than real-time in-person gatherings. At the same time, the chats are equipped with both the speed of instant messaging and the advantage of asynchronous reach (i.e., participants can still catchup on messages sent hours before and provide new life to a discussion).

There are also several reasons for the increase in private messaging usage. For one, these applications enable a measure of encrypted privacy for those who want it — even at the larger scales of group discussion — which is not available on open platforms such as Reddit. For those with data accounts, closed messaging applications also offer a texting, phone, and video service that can work across different countries that either does not require additional cost or is cheaper than an international plan.[5] Whatever the exact reason, the rise of encrypted direct messaging applications is notable; for the last several years, they

---

[3] For example, Simon Kemp, "Digital 2020: Global Digital Overview" (DataReportal in collaboration with HootSuite and WeAreSocial), January 30, 2020, 95, https://datareportal.com/reports/digital-2020-global-digital-overview.

[4] While closed groups in Facebook offer a similar experience to some of the features of encrypted message chat services, and may warrant similar research approaches, we are focusing on the distinguishing characteristics of applications like WhatsApp and Telegram in this paper.

[5] Examples of reasoning can be found in Jordan McMahon, "Why We Should All Ditch Other Messaging Apps for Signal," Wired, November 5, 2017, https://www.wired.com/story/ditch-all-those-other-messaging-apps-heres-why-you-should-use-signal/; Thorin Klosowski, "What's the Deal with All These Messaging Apps?," Lifehacker, April 10, 2014, https://lifehacker.com/whats-the-deal-with-all-these-messaging-apps-1561543034; Meghan Keaney Anderson, "Why We're Thinking About Messaging Apps All Wrong," HubSpot (blog), February 9, 2016, https://blog.hubspot.com/marketing/messaging-apps.

are the only type of mobile application in which first-time and continued use has grown. In 2016, for example, The Economist noted that teenagers spent more time sending instant messages than they did "perusing social networks."[6]

For a sense of proportion when it comes to closed message application usage: at the time of The Economist's reporting in 2016, 1 billion account holders actively utilized Facebook's WhatsApp on a monthly basis. By January 2020, WhatsApp gained an additional 600 million monthly active user accounts or unique monthly visitors, totaling 1.6 billion active users per month; in addition, the application reached 2 billion overall registered accounts in February 2020.[7] Facebook Messenger usage was also very high, totaling 1.3 billion users per month, with WeChat following at 1.1 billion.[8]

These numbers indicate that a significant percentage of Internet users overall are using closed messaging services. In 2018, approximately 49% of the world's population — 3.7 billion Internet users — were reported to use the Internet, a number that rose to 4.1 billion or 53.6 percent of the global population by the end of 2019.[9] Closed messaging applications could include 50% of global Internet users, in other words. Isolation and social distancing resulting from COVID-19 may be contributing to a further spike in global usage: in 2020, not long after the worldwide lockdowns, WhatsApp usage appeared to have increased overall by 51%, with Spain reporting a 76% increase in WhatsApp usage; China's WeChat and Weibo reported increases of 58%.[10]

---

[6] "Mobile Services - Bots, the next Frontier," The Economist, April 9, 2016, sec. Business, https://www.economist.com/business/2016/04/09/bots-the-next-frontier.

[7] Kemp, "Digital 2020: Global Digital Overview"; WhatsApp Blog, "Two Billion Users — Connecting the World Privately," WhatsApp Blog, February 12, 2020, https://blog.whatsapp.com/two-billion-users-connecting-the-world-privately.

[8] Kemp, "Digital 2020: Global Digital Overview."

[9] Susan Halford et al., "Understanding the Production and Circulation of Social Media Data: Towards Methodological Principles and Praxis," New Media & Society 20, no. 9 (September 2018): 3341–58, https://doi.org/10.1177/1461444817748953; International Telecommunication Union, "Statistics," International Telecommunication Union, accessed June 30, 2020, https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

[10] Pranav Hegde, "Coronavirus Impact | WhatsApp Global Usage Spikes by 51% during Pandemic," MSN News: Money Control, March 26, 2020, https://www.msn.com/en-in/news/other/coronavirus-impact-whatsapp-global-usage-spikes-by-51-25-during-pandemic/ar-BB11JFMt; Nic Newman, "Reuters Institute Digital News Report 2020" (Reuters Institute for the Study of Journalism, 2020), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.

The popularity of closed messaging application usage has been high outside of the US, Australia and certain EU countries; regions with dominant closed message usage include South America, Africa, and much of Asia.[11] Despite WhatsApp and WeChat being the most widely used messenger apps worldwide, other competitors have strongholds in regional markets. Examples include Line and KakaoTalk, instant messaging applications from South Korea, or Rakuten-owned Viber. In 2018, 99.2% percent of South Koreans surveyed by the Korean Ministry of Science and Information and Communication Technology were using KakaoTalk.[12] Line was in late 2019 the most popular messaging app in Japan and Taiwan, among other countries.[13]

Given the reach of closed messaging applications, the presence of misinformation or false information within these channels is a cause of grave concern. For example, in December 2020, the Argentinean fact-checking organization Chequeado determined that of the 2,697 debunkings in their database, 778 were identified as circulating within WhatsApp; this amounted to nearly 30% of the misinformation that the organization witnessed circulating during the COVID-19 pandemic.[14]

These concerns are all the more pressing when recognizing that these "closed" applications are inherently porous, with information crossing between them and more "open" spaces within Twitter, YouTube, and Facebook. We already know that people encounter news within these applications increasingly: As of 2018, among people surveyed in 37 countries, the average usage had more than doubled to 16% in four years.[15] The Reuters Institute for the Study of Journalism at Oxford also reported in 2020 that more than half of respondents remain concerned about what is real and fake on the Internet when it comes to online news circulation. In countries such as Kenya and Brazil, where

---

[11] Considering both the top messenger applications in addition to overall global usage, see both We Are Social, "2018 Digital Yearbook," https://www.slideshare.net/wearesocial/2018-digital-yearbook-86862930; Kemp, "Digital 2020: Global Digital Overview."

[12] Korean Ministry of Science and ICT, "2018 년 인터넷이용실태조사 결과 발표 (Announcement of the 2018 Internet Usage Survey)," 대한민국 정책포털 KOREA.KR 정책브리핑 (Korea Policy Portal KOREA.KR Policy Briefing), February 25, 2019, http://www.korea.kr/news/pressReleaseView.do?newsId=156318833.

[13] Kemp, "Digital 2020: Global Digital Overview," 95.

[14] Laura Zommer to Connie Moon Sehat, "[Chequeado Statistics on WhatsApp and Misinformation]," December 23, 2020, https://chequeado.com/latamcoronavirus/.

[15] Nic Newman et al., "Reuters Institute Digital News Report 2018" (Reuters Institute for the Study of Journalism, 2018).

closed messaging apps are dominant, those surveyed correspondingly reported the most concern about misinformation within these social media spaces. The same level of concern exists in Chile, Mexico, Malaysia, and Singapore.[16] While the global prevalence of messenger applications has created a central vehicle for communication among billions of individuals, it has also made flow of false information overall a key worry as well.

Motivations for Analyzing Political Discussions within Closed Spaces

Given their prevalence, closed messaging applications unsurprisingly factor into election environments. Since 2015, scholars have noted that WhatsApp in particular "has established itself as a powerful political tool for spreading political information," whether accurate or not.[17] The earlier 100-member limit within WhatsApp during elections for the Israeli Knesset (Parliament) in 2015, for example, seemed to open possibilities for closer and informal connections with candidates — a potential advantage for political discussion; similar expressions of closeness between political candidates and constituencies were reported in Indonesia as well.[18] After 2016, especially following the revelations of Internet Research Agency- and Cambridge Analytica-powered mass manipulation on Facebook, concerns about disinformation within related social media spaces such as WhatsApp and Facebook Messenger seemed warranted as well.[19]

Given the reach of these tools, it makes sense that they are mobilized for political campaigns around the world.[20] The pivotal

[16] Newman, "Reuters Institute Digital News Report 2020."

[17] Sérgio Barbosa and Stefania Milan, "Do Not Harm in Private Chat Apps: Ethical Issues for Research on and with WhatsApp," Westminster Papers in Communication and Culture 14, no. 1 (August 14, 2019): 49–65, https://doi.org/10.16997/wpcc.313.

[18] Sharon Haleva-Amir, "Not All about That Facebook: Political Campaigns and Civic Engagement in the 2015 Elections," Israel Affairs 22, no. 3/4 (July 2016): 711–26, https://doi.org/10.1080/13537121.2016.1174375; Suwandi Sumartias Sumartono, "Utilization of Whatsapp As A Political Communication Channel Politicians Padang City," in Advances in Social Science, Education and Humanities Research, vol. 136 (2nd International Conference on Social and Political Development, Atlantis Press, 2017), 428–31, https://doi.org/10.2991/icosop-17.2018.67.

[19] Newman et al., "Reuters Institute Digital News Report 2018"; Newman, "Reuters Institute Digital News Report 2020."

[20] Raquel Rennó, "WhatsApp: The Widespread Use of WhatsApp in Political Campaigning in the Global South (Updated Version)," Our Data Our Selves (Tactical Tech) (blog), November 13, 2018, https://ourdataourselves.tacticaltech.org/posts/whatsapp/; Andrew M. Guess and Benjamin A. Lyons, "Misinformation, Disinformation, and Online Propaganda," in *Social Media and Democracy: The State of the*

role of WhatsApp in the 2019 Indian legislative elections revealed how these special spaces could be innovatively used for larger political purposes. WhatsApp Users could be added to closed partisan groups, sometimes unsolicited and without their consent, until April 2019.[21] Tactics included the distribution of misinformation and jokes through local-level WhatsApp initiatives and the organization of WhatsApp groups per polling station around the country.[22] Despite the continuation of in-person rallies, television and newspaper coverage, and traditional canvassing, WhatsApp became the country's default mode of communication and preferred vehicle for campaign messaging.[23] As Ankit Lal, a top strategist for the Aam Aadmi Party, opined: "We wrestle on Twitter. The battle is on Facebook. The war is on WhatsApp."[24]

Yet given the closed nature of this space, it is technologically difficult to explore the nature of these political communications and the extent to which they have a discernible effect upon election outcomes. At the same time, it is unsurprising that journalists, academics, human rights workers, election observers, and other researchers would want to analyze these communication environments better. Reasons include ascertaining the needs and desires of citizens or, more immediately, protecting their interests in the face of potential manipulation via disinformation campaigns. Questions from First Draft for journalists, for instance, make clear the range of interests that news stories might have for such communities. (See Appendix B.)

Field, Prospects for Reform, ed. Joshua A. Tucker and Nathaniel Persily, SSRC Anxieties of Democracy (Cambridge: Cambridge University Press, 2020), 10–33, https://www.cambridge.org/core/books/social-media-and-democracy/misinformation-disinformation-and-online-propaganda/D14406A631AA181839ED896916598500.

[21] Ivan Mehta, "[Best of 2019] WhatsApp Finally Lets You Prevent People from Adding You to Their Shitty Groups," The Next Web (blog), April 3, 2019, https://thenextweb.com/apps/2019/04/03/whatsapp-finally-lets-you-prevent-people-from-adding-you-to-their-shitty-groups/; Chinmayi Arun, "Opinion | India May Be Witnessing the next 'WhatsApp Election' — and the Stakes Couldn't Be Higher," Washington Post, April 25, 2019, https://www.washingtonpost.com/opinions/2019/04/25/india-could-see-next-whatsapp-election-stakes-couldnt-be-higher/; Priyanjana Bengani, "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It," Columbia Journalism Review, October 16, 2019, https://www.cjr.org/tow_center/india-whatsapp-analysis-election-security.php.

[22] Kumar Uttam, "For PM Modi's 2019 Campaign, BJP Readies Its WhatsApp Plan," Hindustan Times, September 29, 2018, https://www.hindustantimes.com/india-news/bjp-plans-a-whatsapp-campaign-for-2019-lok-sabha-election/story-lHQBYbxwXHaChc7Akk6hcI.html.

[23] Vindu Goel, "In India, Facebook's WhatsApp Plays Central Role in Elections," The New York Times, May 16, 2018, sec. Technology, https://www.nytimes.com/2018/05/14/technology/whatsapp-india-elections.html.

[24] Goel.

Specific examples of analysis from 2018 and 2019 elections in Brazil and India illustrate some of the specific concerns of the research community. First, messages from politically oriented groups appear to be quite slanted in these spaces, and they may be deeply interconnected with a wider network of information on open social media platforms. For example, political scientists Camila Mont'Alverne and Isabele Mitozo analyzed a sample of 213 public Brazilian WhatsApp groups with messages containing links to YouTube.[25] Out of these groups, Mont'Alverne and Mitozo found that the largest number of outward-bound YouTube links were for "The Antagonist," a political news media group highly favorable to current President Jair Bolsonaro.[26] In comparison, Bolsonaro's opponents such as Fernando Haddad had visibly fewer mentions in these randomly sampled WhatsApp groups.

Closed messaging communications might in fact also help carry an election. Beyond the anecdotal expressions by WhatsApp participants, the impact of the closed messaging influence appears to have worked in Brazil with the election of Bolsonaro. Bolsonaro — a candidate who had only 8 seconds in traditional political television campaign time — won over 55% of the electoral vote for the presidential election in October 2018.[27] How did this come to pass? Ten days before the 2018 election, Brazil's news publication Folha de São Paulo hinted at a possible answer. Their journalists uncovered a scheme involving millions of campaign messages sent via WhatsApp in favor of Bolsonaro, all of which had been financed by various corporations affiliated with him.[28]

Finally, the misinformation flowing in these chats can foment life-threatening behaviors. Already by 2017, WhatsApp officials in India were becoming concerned about terms of service violations

[25] Camila Mont'Alverne and Isabele Mitozo, "Muito Além da Mamadeira Erótica: As notícias compartilhadas nas redes de apoio a presidenciáveis em grupos de WhatsApp, nas eleições brasileiras de 2018 (Beyond the Erotic Baby Bottle: News shared through WhatsApp by presidential candidates' support networks during the Brazilian elections in 2018)," in Compolítica8 25 (2019), http://ctpol.unb.br/compolitica2019/GT4/gt4_Montalverne_Mitozo.pdf.

[26] Mont'Alverne and Mitozo.

[27] Esther Solano Gallego, "How Bolsonaro Came to Pass," Rosa Luxemburg Stiftung NYC (blog), November 27, 2018, http://www.rosalux-nyc.org/how-bolsonaro-came-to-pass/.

[28] Patrícia Campos Mello, "Empresários bancam campanha contra o PT pelo WhatsApp," Folha de S.Paulo, October 18, 2018, sec. Poder, https://www1.folha.uol.com.br/poder/2018/10/empresarios-bancam-campanha-contra-o-pt-pelo-whatsapp.shtml.

including hate speech, threats of violence, and false statements.[29] False rumors and manipulated media within these chat spaces continued to incite extreme civil unrest, including cases of mob lynching, and contributed to anti-vaccination sentiment.[30] Due to concerns of rampant misinformation, WhatsApp began in 2018 to lower the impact of message reach by introducing limitations on forwards, focusing first on India "where people forward more messages, photos, and videos than any other country in the world."[31] More recently, Telegram served as one location for discussions by members of the far right who called for guns and violence at the U.S. Capitol on January 6, 2021.[32]

The Private-Public Conundrum for Researchers

Closed messaging applications like WhatsApp and Telegram have been used as a political mechanism to influence elections and to undermine the process through violence and civil unrest. Yet these technologies are technically, or technologically, private. Whatever their public impact, closed messaging applications can create mostly private spaces that can thwart hackers and company engineers, in addition to the general public.

Closed Messaging Spaces are Social

This technical point may be a fine one, but it spotlights the larger conundrum: In addition to the designs for privacy, closed messaging spaces have public elements as well. Closed messaging technologies are in fact designed to be *social* — they allow their users to open these spaces up to more public visibility and

[29] Goel, "In India, Facebook's WhatsApp Plays Central Role in Elections."

[30] Soutik Biswas, "Fighting India's WhatsApp Fake News War," BBC News, August 20, 2018, sec. India, https://www.bbc.com/news/world-asia-india-45140158; Cristian Vaccari and Andrew Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," Social Media + Society 6, no. 1 (January 1, 2020), https://doi.org/10.1177/2056305120903408. Referenced in Bengani, "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It."

[31] WhatsApp Blog, "More Changes to Forwarding," WhatsApp Blog, January 21, 2019, https://blog.whatsapp.com/more-changes-to-forwarding.

[32] Craig Timberg and Drew Harwell, "Pro-Trump Forums Erupt with Violent Threats Ahead of Wednesday's Rally against the 2020 Election," Washington Post, January 5, 2021, https://www.washingtonpost.com/technology/2021/01/05/parler-telegram-violence-dc-protests/.

impact. We should acknowledge that the line between public and private conversation within these applications is explicitly blurred, some less than others.

Differently than WhatsApp, Telegram allows users to define a public group, which is known to be discoverable in search and available to anyone who wants to join.[33] But consider the scale of any of these groups, publicized or not. Message spaces with a default technological capacity for group membership of 256 members in the popular instance of WhatsApp, 1,000 in the case of Signal, or even 100,000 for Telegram defy any reasonable expectation of intimate conversation.[34]

Furthermore, group administrators and users sometimes publish invitations to chat groups through mechanisms that permit open sharing: invitations to groups can be posted as a URL on websites, or in Instagram and Pinterest postings, for example. As another example of the public nature of closed chat spaces, Viber paired with an Israeli news station in 2015 to host a moderated public discussion related to the Knesset elections.[35] Should a group decide to post an invitation to a group publicly, it seems logical to assume that the conversation may no longer be considered private.

Allowing for features such as ease of sharing or discoverability reduces some technical security and safety features. On WhatsApp, media and other attachments are encrypted separately from the main message, and a pointer to the encrypted media item is sent along with the message to enable ease of sharing of heavy files.[36] On Signal, when a person joins the app, all their contacts are notified of their joining the platform. This is a feature aimed at increasing social interactions and the user base of the app, but it restricts people who might want to use the platform secretly.[37] Telegram only offers end-to-end encryption

---

[33] "Supergroups 5000: Public Groups, Pinned Posts," Telegram, March 13, 2016, https://telegram.org/blog/supergroups5k.

[34] "Group Chats," Signal Support, accessed February 5, 2021, https://support.signal.org/hc/en-us/articles/360007319331-Group-chats.

[35] Haleva-Amir, "Not All about That Facebook."

[36] WhatsApp, "WhatsApp Encryption Overview: Technical White Paper [Version 3]," October 22, 2020, https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf.

[37] Billy Perrigo, "The Inside Story of How Signal Became the Private Messaging App for an Age of Fear and Distrust (Updated)," Time, September 28, 2020, https://time.com/5893114/signal-app-privacy/.

within the "secret chat" protocol, which must be explicitly chosen.[38]

Yet Companies Communicate and Users Expect Privacy

This private-public hybridization creates awkwardness for closed messaging technology from a policy perspective. For example, amid COVID-19 misinformation concerns in 2020, WhatsApp has continued to emphasize its efforts to "constrain virality" and to signal "less personal" forwarded messages through the double arrow icon. The threat, at least according to WhatsApp measures, appears to be an external rather than internal one. Thus, the goal of the application's builders is to keep the space "personal," "private," and "secure" — both in terms of encryption and of misinformation.[39]

Ownership of these messages is explicitly granted to end users, rather than the public domain. This understanding is supported by application terms of service. According to WhatsApp's terms of service: "Your messages are yours, and we can't read them. We've built privacy, end-to-end encryption, and other security features into WhatsApp. We don't store your messages once they've been delivered. When they are end-to-end encrypted, we and third parties can't read them."[40] Contrast this with the Telegram privacy policy, which emphasizes a bit differently that they "don't use your data to show you ads," since data belongs to the end user. Private chats are, as mentioned earlier, differentiated from public ones in their policy and tool.[41]

Further contributing to this sense of privacy is the assurance that application owners and developers cannot read the messages — although they are privy to quite a lot of metadata — because of the encryption mechanisms. To study what messages are flowing through these channels, analysts and researchers must join these private chat groups in ways that raise questions regarding the boundaries of ethical research in these contexts.

---

[38] "Telegram FAQ," Telegram, accessed January 11, 2021, https://telegram.org/faq#q-how-secure-is-telegram.

[39] WhatsApp Blog, "Keeping WhatsApp Personal and Private," WhatsApp Blog, April 7, 2020, https://blog.whatsapp.com/Keeping-WhatsApp-Personal-and-Private.

[40] "WhatsApp Legal Info," WhatsApp.com, accessed July 8, 2020, https://www.whatsapp.com/legal/#key-updates.

[41] See "Telegram FAQ" and Telegram, "Telegram Privacy Policy," Telegram, accessed February 19, 2021, https://telegram.org/privacy.

The technology's hybrid nature, such as the public availability of chat links, does not completely resolve the question of user expectations. In even ostensibly public arenas like Twitter, users have been found to express discomfort about research of their tweets without their consent, let alone surprise that the researchers are studying their messages.[42]

Related to closed spaces, users have expressed surprise that their groups were even discoverable. Journalists reported in February 2020 that Google Search had been indexing WhatsApp Group invite links and rendering them publicly discoverable. Initial responses from both platforms were unsurprised; after all, this is what the technology has been built to accomplish. Facebook pointed out: "Like all content that is shared in searchable public channels, invite links that are posted publicly on the internet can be found by other WhatsApp users." A Google representative responded in similar fashion:

"[S]earch engines like Google & others list pages from the open web. That's what's happening here. It's no different than any case where a site allows URLs to be publicly listed."[43]

Despite these replies, Google took steps within a few days to remove the indexing, while WhatsApp supplied a no-follow rule to discourage automatic indexing as well.[44] What was key, as demonstrated by news coverage on the topic, was the user's sensibility of private and public discussion. "WhatsApp groups may not be as secure as you think they are," warned the Deutsche Welle reporter who initially broke the story.[45]

The technology, which is a bit confused when it comes to the public/private divide, therefore presents a conundrum for researchers overall. Already, in explicitly public social media

---

[42] Casey Fiesler and Nicholas Proferes, "'Participant' Perceptions of Twitter Research Ethics," *Social Media + Society* 4, no. 1 (January 1, 2018): 2056305118763366, https://doi.org/10.1177/2056305118763366.

[43] Kim Lyons, "Google Is Indexing WhatsApp Group Chat Links, Making Even Private Groups Discoverable," The Verge, February 21, 2020, https://www.theverge.com/2020/2/21/21147073/whatsapp-google-group-chat-join-indexing-links-search-privacy-facebook.

[44] Lisa Vaas, "Google Stops Indexing WhatsApp Chats; Other Search Engines Still at It," Naked Security (blog), February 25, 2020, https://nakedsecurity.sophos.com/2020/02/25/google-stops-indexing-whatsapp-chats-other-search-engines-still-at-it/.

[45] Joseph Cox, "VICE - Google Is Letting People Find Invites to Some Private WhatsApp Groups," Motherboard (Vice), February 21, 2020, https://www.vice.com/en_us/article/k7enqn/google-is-letting-people-find-invites-to-some-private-whatsapp-groups.

arenas like Twitter, ethical norms for research are not clear. Closed messaging applications are technologies with large public effects as well, but they take place within applications that are designed to be mostly private, and this confusion poses additional difficulty for professional ethics.

Models of Research Investigation

Overall, many of the studies reviewed did take steps to observe the growing practice of protecting individual privacy, for example either omitting or de-identifying phone number accounts within data collections.[46] In one case, a study revealed phone numbers only at the country code level.[47] In addition, some research efforts reported special data storage handling, such as the encryption of text messages or only using local hard drives.[48] We return to some of these practices in Appendix A.

How to protect subject identities is only the beginning of the choices that researchers have to make when it comes to their ethical obligations. Researchers have approached the tension between public and private expectations in chat spaces in different ways. To clarify the considerations that researchers try to navigate, our review has narrowed the scope of examination in two ways: on projects primarily focused around democratic elections and upon systematized collections of message content.

Elections in democratic contexts have specific frameworks of expectation related to both human and citizenship rights, which include freedom of opinion and expression, transparency and access to information.[49] The collection of messages also touches on the right to privacy, which is a civil liberty protected in the

[46] One example of the transformation to the practice can be seen in the establishment of the US Federal Policy for the Protection of Human Subjects known as the "Common Rule" which was updated in 2018 to reflect digital data collection practices, see U.S. Department of Health and Human Services, "Federal Policy for the Protection of Human Subjects ('Common Rule')" (2018), https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

[47] See for example Kiran Garimella and Gareth Tyson, "WhatsApp, Doc? A First Look at WhatsApp Public Group Data," in Proceedings of the Twelfth International AAAI Conference on Web and Social Media (International AAAI Conference on Web and Social Media, Stanford, CA: AAAI Press, 2018), https://aaai.org/Library/ICWSM/icwsm18contents.php; Josemar Alves Caetano et al., "Analyzing and Characterizing Political Discussions in WhatsApp Public Groups," ArXiv:1804.00397 [Cs], April 2, 2018, http://arxiv.org/abs/1804.00397.

[48] See Garimella and Tyson, "WhatsApp, Doc?"; Bengani, "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It."

[49] The Carter Center hosts a database of over 200 sources of public international law on the topic at https://eos.cartercenter.org/.

Universal Declaration of Human Rights as well as many democratic legal traditions.[50]

In addition, there have been increasing examples of studies that collect these messages systematically. As a baseline, researchers are well practiced at gathering data for analysis in research conditions under explicit informed consent. Of studies aspiring to newer methodologies around larger scales of data collection, however, four models of research appear to be in operation — voluntary contribution, collection through focused partnerships, entrance without disclosing researcher intent, and entrance with disclosure — which are explained further below.[51] They all define the public nature of their research, though in different ways.

Baseline Model: Smaller Investigations under Informed Consent

When focusing on studies that examine message texts, research studies can sidestep some of the thornier privacy and human subject issues through design, by focusing only on what can be shared through an umbrella of informed consent.

Traditional methods of qualitative field study, specifically surveys and interviewing, continue to offer an avenue that can help characterize political behaviors around an election — such as generational trends in Spain, the United States, and New Zealand or other aspects of community in Sierra Leone —without divulging specific examples of messages.[52] Mixed-methods

[50] United Nations General Assembly, "Universal Declaration of Human Rights (UDHR)," December 10, 1948, https://www.un.org/en/universal-declaration-human-rights/index.html.

[51] Examples of studies can be found in Bruno Schaefer et al., "Qual o Impacto Do WhatsApp Em Eleições? Uma Revisão Sistemática (2010-2019)," Revista Debates 13, no. 3 (December 16, 2019): 58–88, https://doi.org/10.22456/1982-5269.96255.

[52] Maria Frahm-Arp, "The Political Rhetoric in Sermons and Select Social Media in Three Pentecostal Charismatic Evangelical Churches Leading up to the 2014 South African Election," Journal for the Study of Religion 28, no. 1 (2015): 115–41; Homero Gil de Zúñiga, Alberto Ardèvol-Abreu, and Andreu Casero-Ripollés, "WhatsApp Political Discussion, Conventional Participation and Activism: Exploring Direct, Indirect and Generational Effects," Information Communication and Society, July 20, 2019, 1–18, https://doi.org/10.1080/1369118X.2019.1642933; Maggie Dwyer, Jamie Hitchen, and Thomas Molony, "Between Excitement and Scepticism: The Role of WhatsApp in Sierra Leone 2018 Elections," in *Social Media and Politics in Africa: Democracy, Censorship and Security*, ed. Maggie Dwyer and Thomas Molony (Zed Books Ltd., 2019); Nic Cheeseman et al., "Social Media Disruption: Nigeria's WhatsApp Politics," Journal of Democracy 31, no. 3 (July 14, 2020): 145–59, https://doi.org/10.1353/jod.2020.0037; Edson C Tandoc, Darren Lim, and Rich Ling, "Diffusion of Disinformation: How Social Media Users Respond to Fake News and Why," Journalism 21, no. 3 (March 2020): 381–98, https://doi.org/10.1177/1464884919868325.

approaches may also include the analysis of app metadata in combination with participant questionnaires.

These methods may rely on the ability of participants to answer questions honestly and accurately. They also, as with larger quantitative studies, rely on the capacity of investigators to appropriately contextualize the results.

Model 1: Voluntary Contributions

When it comes to projects that collect message data from closed messaging spaces at a large scale, one model relies on voluntary contributions from application users. This allows researchers not to join the groups and enables an exchange that indicates consent. Two examples of the contribution model include a "tip line" approach, where users send in their misinformation tips to a central number, and broadcast messages.

Two examples show how the tip line approach has worked in different ways; through their advertising of national WhatsApp numbers, the attempted scale of these examples is demonstrably different from smaller studies of forwarded messages among a fixed set of participants.[53]

- The Comprova project during the 2018 elections in Brazil consisted of a coalition of 24 Brazilian newsrooms, backed by the First Draft Network. The coalition received tips of potential misinformation on WhatsApp, including message texts and images, that an assigned journalist would then research and report back to the submitter as well as to the public at large.[54] Any member of a WhatsApp group could voluntarily submit a fact-check request within a conversation without revealing the identities of other group members or other aspects of the discussion.

[53] Irene V. Pasquetto et al., "Understanding Misinformation on Mobile Instant Messengers (MIMs) in Developing Countries" (Harvard Shorenstein Center on Media, Politics and Public Policy, May 2020).

[54] Anna Jean Kaiser, "The Brazilian Group Scanning WhatsApp for Disinformation in Run-up to Elections," The Guardian, September 26, 2018, sec. World news, https://www.theguardian.com/world/2018/sep/26/brazil-elections-comprova-project-misiniformation-whatsapp; Claire Wardle et al., "Comprova: An Evaluation of the Impact of a Collaborative Journalism Project on Brazilian Journalists and Audiences" (First Draft, June 27, 2019), https://firstdraftnews.org:443/latest/comprova-an-evaluation-of-the-impact-of-a-collaborative-journalism-project-on-brazilian-journalists-and-audiences/.

- WhatsApp directly commissioned a similar model of tip line investigation during the 2019 Indian elections through the Checkpoint project, led by the Proto enterprise in partnership with Dig Deeper Media and Meedan companies. An important difference in this project model was that it was oriented more toward research for analysts than intended as a helpline for the public, which meant that not every submission received a response.[55]

As opposed to analyzing messages collected through tip lines, or many voluntary submissions to one single number (M:1 ratio), the reverse is also possible. The collection of messages for one-to-many broadcasts (1:M), with the researcher account(s) as receivers of broadcast information, allows for a different kind of analysis which could potentially scale over time.

- Researchers during the 2015 presidential elections in Nigeria analyzed WhatsApp broadcast messages. WhatsApp broadcasts are sent from a sender to recipients on a defined list but can only be received if the recipient has stored the sender's phone number as well, presumably to prevent spamming.[56] By analyzing the content sent by select accounts to many recipients —a mechanism of implicit consent through the mutual storage of phone numbers —the researchers were able to give a sense of the tone, accusations, and assertions around political candidates that were part of the election media environment without entering into private chat situations or seeing others' responses.

Model 2: Join Groups through Invitation with a Public/Publicized Purpose

Another model for evaluating messages, but this time with members of the analysis team entering directly into chat spaces

---

[55] Abhimanyu Ghoshal, "WhatsApp Launches a Tip Line in India to Battle Fake News," The Next Web, April 2, 2019, https://thenextweb.com/apps/2019/04/02/whatsapp-launches-a-tip-line-in-india-to-battle-fake-news-ahead-of-national-elections/; Ryan Mac and Pranav Dixit, "The Tip Line WhatsApp Launched to Combat Fake News Isn't Actually Going to Combat Fake News," BuzzFeed News, April 3, 2019, https://www.buzzfeednews.com/article/ryanmac/whatsapp-fake-news-tip-line-indian-election-not-helpline; "PROTO | Checkpoint 'FAQs About Project Checkpoint,'" Proto, accessed July 8, 2020, https://www.checkpoint.pro.to/.

[56] Oluwabunmi O. Oyebode and Adeyemi Adegoju, "Appraisal Resources in Select WhatsApp Political Broadcast Messages in the 2015 Presidential Election Campaign in Nigeria," Journal of Pan African Studies 10, no. 10 (November 1, 2017): 29; "WhatsApp FAQ — How to Use Broadcast Lists," WhatsApp.com, accessed July 13, 2020, https://faq.whatsapp.com/iphone/chats/how-to-use-broadcast-lists.

themselves, was established through a collaborative election tracking project.

- During the Ghana 2016 general election, a collaboration of organizations established a "Social Media Tracking Center" (SMTC) to prevent violence and other election threats by monitoring messages on social media.[57] The tracking center model and software platform had previously been established in other country election contexts to examine Twitter and Facebook information but also, for the Ghana election, included the collection of WhatsApp messages. In advance of the election, members of Pen Plus Bytes (a local civil society organization), United Nations University, Ghana's National Election Security Task Force, the Electoral Commission and others established protocols of communication for verified incidents.

  A single SMTC WhatsApp account, which was publicly shared, meant that citizens could submit their own questions and reports. In addition, this account was permitted access by group administrators to a set of WhatsApp groups for incident tracking. The texts of all messages were collected — something explained to the members of the WhatsApp group by Pen Plus Bytes and the research team. This fuller collection allowed the SMTC software to automatically flag potential issues for further examination by a team of human reviewers who could escalate issues when necessary.[58]

Unlike Model 1, this project collected all messages from the closed group for a certain period, which meant that analysts could examine both the text of a message and the account details of the message sender, in addition to the conversational context. At the same time, the purpose of the project was public and publicized.

---

[57] Andrés Moreno, Philip Garrison, and Karthik Bhat, "WhatsApp for Monitoring and Response during Critical Events: Aggie in the Ghana 2016 Election," in Proceedings of the 14th ISCRAM Conference (ISCRAM 2017, Albi, France, 2017), 11.

[58] According to the project, there was an "introductory text message that was sent [by members of Pen Plus Bytes and the research team] from the SMTC phone to the WhatsApp groups being monitored. It briefly explained what the SMTC was and that all the messages in the group were being scooped up," Philip Garrison to Connie Moon Sehat, "[Request for Clarification] Research Practices for Closed Messaging Groups," July 22, 2020.

Model 3: Join "Public" Groups Without Disclosing Research Intent

The third model leverages the ambiguity of private chat invitations, or links available on the open web, to enter closed groups and study them. In addition, researchers did not announce their presence or purpose in joining the chats. A few accounts included an identification of a research institute in their profile data. However, posting one's affiliation assumes that chat participants are able to identify research institutes that are outside their countries and are comfortable querying other members of a group about their presence (e.g., perhaps the person is a researcher but not in the group as an observer). This is not the same as a formal disclosure of researcher presence for the purpose of analysis.

This model raises the question of whether the notion of a "public group" — when publicized and made available for general entry — exists among the public or perhaps for a specific context that lowers the barrier of entry for research purposes.

Two studies show how researchers might identify themselves differently:

- Researchers can use the "name" field of the account to identify themselves, but not make declarations beyond this. Kiran Garimella and Gareth Tyson designed a methodology around a software process to scrape what they defined as public WhatsApp data at a larger scale than typical studies. The authors took chat messages from a sample of 200 groups from 2,500 WhatsApp groups in 2018 that were publicly indexed.[59] When joining such groups, the researchers identified the accounts as being "researchers from" an academic organization. There was, however, no announcement within the group chat about their joining or presence as researchers.[60] This

[59] Garimella and Tyson, "WhatsApp, Doc?"

[60] Kiran Garimella to Connie Moon Sehat, "Re: [Air-L] [Request for Input] Research Practices for Closed Messaging Groups," March 25, 2020.

methodology and process has now been systematized further by the WhatsApp Monitor system project.[61]

- Another example of this approach —identification but no announcement — took place during the 2019 India elections.[62] The collection of over 1 million messages by the Tow Center for Journalism at Columbia University for the period between April 11 and May 19, 2019, leveraged "groups of which we were already a part" as well as other groups with publicized invite links of at least 60 members. The study noted that they were at times banned from WhatsApp or dropped from groups but that the rationale for removal was not clear (e.g., automated group joins from WhatsApp Desktop potentially flagged as bots, a foreign-based account number flagged to administrators).[63]

Examples of research projects that have collected data without any indication of a research purpose or identity:

- Leveraging the public group feature of Telegram, the Pushshift Telegram dataset as of mid-2020 allows researchers to evaluate 317 million messages, 2.2 million unique users, and 27,800 thousand channels.[64] The collection comprises messages and metadata from the Washington Post's public chat group. Examples of collected information, in addition to the actual messages sent in the channel, also include "message data exposed by the Telegram API … the datetime that the message was sent, whether it included media (e.g. images or video), and the identification number of the user who sent the message. For each message sender, Telegram provides details such as their username, whether they are a bot, whether they are a verified user, and so on."[65]

---

[61] Philipe Melo et al., "WhatsApp Monitor: A Fact-Checking System for WhatsApp," Proceedings of the International AAAI Conference on Web and Social Media 13 (July 6, 2019): 676–77.

[62] "[T]o be fully transparent, we identified ourselves as 'Tow Center' with the description 'a research group based out of Columbia University in New York,'" from Priyanjana Bengani to Aleksei Kaminski, "Email Reply: Introducing Myself and Your Research," March 18, 2020.

[63] Bengani, "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It."

[64] Jason Baumgartner et al., "The Pushshift Telegram Dataset," Proceedings of the International AAAI Conference on Web and Social Media 14 (May 26, 2020): 840–47.

[65] Ibid.

- In the Mont'Alverne and Mitozo example from 2018 in Brazil cited earlier, the researchers did not share their purposes.[66] According to one of the researchers, they felt that divulging their presence would change the dynamics of the conversation and would likely result in further exclusions from groups; however, they stressed a protocol that strove to preserve identities in the research. Having collected their WhatsApp invitation links from open social networking locations on Twitter and Facebook in addition to Google, the researchers also stressed that they considered these to be public arenas and open for entry.[67]

Other cases from Brazil in 2018 and 2019 from researchers at the Pontifícia Universidade Católica de Minas Gerais (PUC Minas) and Universidade Federal de Minas Gerais (UFMG) – in addition to IBM Research and the Max Planck Institute for Software Systems (MPI-SWS) in Germany —also appear in part to have followed the same protocol of not indicating researcher presence.[68]

A joint project among researchers from institutions with different requirements can result in a mix of named and non-named accounts. For example, for a project across MIT, MPI-SWS, and other institutions, the part of the data collection that took place under MPI-SWS did not have to reflect the research institution in the accounts that collected data.[69]

In contrast to most of the cases from the first models, research results from Model 3 were not directly communicated back to

---

[66] Mont'Alverne and Mitozo, "Muito Além da Mamadeira Erótica: As notícias compartilhadas nas redes de apoio a presidenciáveis em grupos de WhatsApp, nas eleições brasileiras de 2018 (Beyond the Erotic Baby Bottle: News shared through WhatsApp by presidential candidates' support networks during the Brazilian elections in 2018)."

[67] « Nós fizemos a opção de não nos identificar ao entrar nos grupos. Isso aconteceu porque não haveria outra forma de realizar a pesquisa. Os grupos tendem a ser muito fechados e nos identificar como pesquisadores alteraria completamente a dinâmica e, em outros casos, seríamos excluídos (já fomos excluídos de alguns sem mesmo nos identificar), » Camila Mont'Alverne to Aleksei Kaminski, "E-Mail Reply: Introdução e Sua Pesquisa-Aleksei Kaminski," March 18, 2020; Camila Mont'Alverne to Aleksei Kaminski and Connie Moon Sehat, "Re: Oi e Conselho! - Aleksei," July 23, 2020.

[68] Other examples include Caetano et al., "Analyzing and Characterizing Political Discussions in WhatsApp Public Groups"; Gustavo Resende et al., "Analyzing Textual (Mis)Information Shared in WhatsApp Groups," in *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19 (Boston, Massachusetts, USA: Association for Computing Machinery, 2019), 225–34, https://doi.org/10.1145/3292522.3326029; Gustavo Resende et al., "(Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures," in *The World Wide Web Conference on - WWW '19* (The World Wide Web Conference, San Francisco, CA, USA: ACM Press, 2019), 818–28, https://doi.org/10.1145/3308558.3313688.

[69] Fabricio Benevenuto to Aleksei Kaminski and Connie Moon Sehat, "Introducing Myself and Your Research- Aleksei," September 3, 2020.

chat group members or to civil society and government officials; rather, their output was for academic purposes of understanding. Their research designs, sometimes approved by university review boards, may not have required the researchers to report their findings back to the members of groups they joined.

Model 4: Join 'Public' Groups while Disclosing Research Intent

A final example demonstrates how researchers have leveraged the ambiguity of the publicly indexed chat groups while disclosing their researcher identities to group members.

- In a different case from the Indian 2019 elections, researchers from Oxford University identified themselves in two ways. After finding links to chat groups index on the web that fit their parameters, they announced their identity upon gaining entry to the chat space. The purpose was to offer "group members the option to withdraw consent from participating in this study. This resulted in some administrators removing us."[70]

This study also opted not to analyze any user information or personal communications within messages. Instead, the project focused on "public" topical content shared, such as URLs or visual media considered to be part of the public domain.[71]

Design Questions and Choices for Researchers

It may seem obvious, but all of these studies discussed and cited are clearly operating under the assumption the chat groups under

[70] Vidya Narayanan et al., "News and Information over Facebook and WhatsApp during the Indian Election Campaign," Computational Propaganda Project Data Memo (Oxford, UK: Oxford Internet Institute, May 13, 2019), https://comprop.oii.ox.ac.uk/research/india-election-memo/ (now at https://comprop.oii.ox.ac.uk/research/posts/news-and-information-over-facebook-and-whatsapp-during-the-indian-election-campaign/); Vidya Narayanan to Connie Moon Sehat, "Re: [Request for Input] Research Practices for Closed Messaging Groups," July 28, 2020.

[71] Discussed as an example in this paper through Narayanan et al., "News and Information over Facebook and WhatsApp during the Indian Election Campaign"; Earlier example in Caio Machado et al., "A Study of Misinformation in WhatsApp Groups with a Focus on the Brazilian Presidential Elections.," in *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19 (New York, NY, USA: Association for Computing Machinery, 2019), 1013–19, https://doi.org/10.1145/3308560.3316738.

analysis are forums of public exchange, else, these research projects would not be publicized. Yet they differ in technologies, size, topics, and scope.

To help clarify the implicit or explicit decisions that researchers are making, we have identified six questions that appear central, drawing from earlier considerations of related topics and across workshops held in collaboration with the MisinfoCon conference series.[72] We explain these questions and implications in depth to help researchers with greater clarity about the ethical validity of embarking upon research projects around closed message texts in democratic contexts.

The answers to these questions can be used to examine research designs. In addition, answers to these questions can be taken in combination with practical suggestions for data collection, available in Appendix A.

One overall practical suggestion for researchers is to submit their research protocols to an external trusted colleague for review. Whether or not researchers submit their protocols to Independent Review Boards, projects involving research in the public interest can benefit from peer feedback related to security and obligations.

Public parameters. Questions regarding the nature of the research effort are asked in tandem:

- Under which conditions is a closed message chat public?
- Who does the data belong to?

Security requirements. In addition, researchers need to think about their own capacity to conduct the research in a secure fashion:

- What is the plan for storing closed messaging research data in a secure fashion?

[72] For example, Connie Moon Sehat, "Closed Messaging Spaces: Thinking About Best Practices and Ethics in a Global Environment," Medium (blog), October 4, 2018, https://misinfocon.com/closed-messaging-spaces-thinking-about-best-practices-and-ethics-in-a-global-environment-10fc82191743.

- How might the research project affect group members and their conversations?

Researcher obligations. Final questions deal with the obligations of the researcher, once the public nature of the group chat is established:

- What are the obligations for disclosure and informed consent?
- When should researchers inform or report back to the groups involved in the findings of their studies?

Additionally, Appendix A offers suggestions of scenario planning that may be helpful in operationalizing the above questions.

Parameters: Exactly When Is a Closed Message Chat 'Public'?

What are the conditions in terms of indexed invites, group size, discussion topics, or other aspects of closed message discussions that make them arguably public versus private?

Technological Signals: Publicly Indexed or Available Invitations, Group classification

While it remains unclear how pressing this consideration is given February 2020's removal of WhatsApp URL indexing within Google search, the question remains: Does a publicized invitation render a closed message space conversation "public"? In the European context, for example, the notion of the "right to respect for his or her family life, home and communications" is strong, so understanding this boundary is important.[73]

In the case of Telegram, an additional possibility includes the explicit choice of making a group public.

Public Interest: Topic and Legal Considerations

First Draft in 2018 explained that its tip line approach seemed to be the only way to honor the terms of service, at least within

[73] "Charter of Fundamental Rights of the European Union," Article 7, Official Journal of the European Union, no. C 83 (March 30, 2010): 389–403.

certain jurisdictions: "Due to WhatsApp's end-to-end encryption, this method of soliciting tips is the only possible way to collect misinformation data without violating the app's terms of service."[74]

At the same time, the need to understand the broad impact of social media technologies is contributing to new interpretations of "fair use" in service of the public interest. In the United States, recent rulings have sanctioned terms of service violations when researching bias and racial discrimination in online hiring websites. The creation of false logins or identities, a breach of website terms of service, has recently been judged not to be a violation of the Computer Fraud and Abuse Act for certain research purposes.[75] Upcoming cases will also continue to wrestle with the legality of web scraping on arguably public, or at least, publicly available, LinkedIn profile pages.[76] Additionally, the European Union's Code of Practice on Disinformation indicates willingness by signatories to:

[S]upport good faith independent efforts to track [d]isinformation and understand its impact, including the independent network of fact- checkers facilitated by the European Commission upon its establishment. This will include sharing privacy[-]protected datasets, undertaking joint research, or otherwise partnering with academics and civil society organizations if relevant and possible.[77]

Signatories include platforms such as Google, Facebook, Microsoft, Twitter, and TikTok.[78]

In less controversial settings, consider evolutions of the explicitly public gatherings in WhatsApp spaces by governmental authorities. More recently than the earlier cited Israeli Knesset example, the model of the Nakuru Analysts in Kenya demonstrates the bringing together of elected county officials and

---

[74] Pedro Burgos, "What 100,000 WhatsApp Messages Reveal about Misinformation in Brazil," First Draft (blog), June 27, 2019, https://firstdraftnews.org:443/latest/what-100000-whatsapp-messages-reveal-about-misinformation-in-brazil/.

[75] Sandvig v. Barr, No. 16–1368 (D.D.C. March 27, 2020).

[76] hiQ Labs, Inc. v. LinkedIn Corporation, No. 17–16783 (9th Cir. September 9, 2019).

[77] European Union, "European Union Code of Practice on Disinformation" (European Union, September 26, 2018).

[78] Ibid.

citizens in one space for the purpose of governance.[79] Freetown City Council of Sierra Leone recently announced its intent to create digital town halls using WhatsApp as well.[80]

Group Size and Expectation

Finally, size also factors into a definition of public conversation. The expectation of privacy, or at least a private, intimate conversation, among a 100,000- or even 200-member chat may be low, no matter the topic. It may be that a group conversation of five or 10, even if with a publicly indexed link on a topic of public interest, still has an appropriate expectation of privacy. One participant in our workshops offered the perspective that even a closed chat of three individuals might be considered "public"; given the potential for copies of messages shared and leaked as the number of group participants increases, this position is perhaps closer to how users should modify their expectations about closed messaging platforms. The Columbia study used a threshold of 60 members as crossing over into the realm of public analysis, though it should be noted that the threshold can only be determined after a researcher joins the group.

It bears consideration that efforts are critically needed that educate users on reasonable expectations of privacy in large group chats, perhaps to be taken up by all technology platforms, researchers, and civil society organizations.

Overlapping Methodological Definitions of Public and Private

Clearly, there are a number of approaches to defining what constitutes a public and private space. Given the blended nature of the closed messaging space, it seems helpful for researchers to clarify the methodological traditions upon which they rely and where there are overlaps with other considerations.

For example, consider sociological and technical interpretations of terms such as "private" and "encrypted." To what extend do they overlap and how do they differ? The term "intimate" and

---

[79] Duncan Omanga, "WhatsApp as 'Digital Publics': The Nakuru Analysts and the Evolution of Participation in County Governance in Kenya," Journal of Eastern African Studies 13, no. 1 (January 2, 2019): 175–91, https://doi.org/10.1080/17531055.2018.1548211.

[80] Freetown City Council, "FCC to Start Digital Town Hall Initiative — Freetown City Council," Freetown City Council (blog), September 24, 2020, https://web.archive.org/web/20201021133638/https://fcc.gov.sl/fcc-to-start-digital-town-hall-initiative/.

"personal" might also be useful from a sociological perspective versus a psychological one.

In addition, specific country contexts will require a consideration of political and legal conceptions of public and private, especially when these conceptions are being challenged. Recent developments in The Gambia, Nigeria, and Kenya, which involve proposed social media legislation or the right of the government to surveil speech, are examples of how research definitions might need to be adjusted to the context.[81]

Parameters: Who Does the Data Belong To?

Another issue that is implicated in the establishment of a study's ethical validity has to do with the ownership of data in closed messaging spaces. Defining who owns the data has implications for how research is conducted and how the data is stored or shared. As discussed earlier, message texts belong to the message writer, at least according to two examples of terms of service. However, decisions such as the *Sandvig v. Barr* decision in the United States indicates an increasing acceptance of such research under certain conditions and in the public interest.

Furthermore, from the Oxford Internet Institute example of message analysis (under Model 4), we see a practice in which the researchers employed a content-oriented strategy. They focused not on personal texts but generally accessible media such as URLs and images shared in the information network. Perhaps in this case, if focusing on a particular element of data, the considerations are different.

Data related to the message writers and receivers, such as phone numbers, also belong to the account holders. In aggregate, however, to whom does the data belong, and in what cases may it be shared? The European Union's General Data Protection

[81] Idayat Hassan and Jamie Hitchen, "Forums of Debate? WhatsApp and the Gambia's Political Transition" (Centre for Democracy and Development, n.d.), originally at https://cddelibrary.org/wp-content/uploads/2020/09/FORUMS-OF-DEBATE-WHATSAPP-AND-THE-GAMBIAS-POLITICAL-TRANSITION.pdf, now at https://docplayer.net/195505832-Forums-of-debate-whatsapp-and-the-gambia-s-political-transition-by-idayat-hassan-and-jamie-hitchen-september-2020.html; "The Gambia," Freedom on the Net 2020 (Freedom House, 2020), https://freedomhouse.org/country/gambia/freedom-net/2020; Communications Team, "Digital Rights in Africa Report 2019," Paradigm Initiative, accessed December 3, 2020, https://paradigmhq.org/download/dra19/; Emmanuel Paul, "Everything You Need to Know about Nigeria's Social Media Bill and What You Can Do about It," Techpoint Africa (blog), November 28, 2019, https://techpoint.africa/2019/11/28/nigerias-social-media-bill/.

Regulation has the strongest considerations of what is considered personal data, which at times includes the IP addresses through which citizens are active.[82] Yet when a citizen voluntarily enters a public arena, some identification is necessarily divulged; defining the context for communication is important. In the end, whether public or private, data collected for the purposes of the public (or the public interest) requires careful handling.[83]

Security: What is the Plan for Storing Closed Messaging Research Data in a Secure Fashion?

There is an overall researcher responsibility when it comes to automated processes and data collection, since choices related to data "can have very real implications to doing right or wrong, even if, *prima facie*, it seems merely to be a morally neutral matter of choosing one's tools."[84] Especially related to obligations to de-identify or anonymize personally identifiable information, a good question for researchers to ask is whether they have the capacity to implement secure data collection. If teams do not possess the ability to protect and steward the data, designing a project that does not store this information is recommended.

Security: How Might the Research Project Affect Group Members and Their Conversations?

Since Werner Heisenberg's "uncertainty principle" was first articulated in 1927, many researchers have applied the recognition that scientific observation will affect the actual interactions observed far beyond the realm of quantum mechanics.[85] In moving ahead with research projects, in addition to recognizing that observational data is always imperfect, the question remains

---

[82] For example, European Union General Data Protection Regulation, Article 3, European Union, "General Data Protection Regulation (GDPR)," REGULATION (EU) 2016/ 679 § Art. 4 GDPR - Definitions (2018), https://gdpr.eu/article-4-definitions/.

[83] Philip M. Napoli, "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers," Telecommunications Policy (Special Issue on the Governance of Social Media), 39, no. 9 (October 1, 2015): 751–60, https://doi.org/10.1016/j.telpol.2014.12.003.

[84] Miles C. Coleman as quoted in Markus Wiesenberg and Ralph Tench, "Deep Strategic Mediatization: Organizational Leaders' Knowledge and Usage of Social Bots in an Era of Disinformation," International Journal of Information Management 51 (November 26, 2019), https://doi.org/10.1016/j.ijinfomgt.2019.102042; Meg Leta Jones, "Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication," Communication Law and Policy 23, no. 2 (April 3, 2018): 159–95, https://doi.org/10.1080/10811680.2018.1430418.

[85] Noting that Heisenberg himself did not use the term "principle," the implications are the same, see Jan Hilgevoord and Jos Uffink, "The Uncertainty Principle," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2016 (Metaphysics Research Lab, Stanford University, 2016), https://plato.stanford.edu/archives/win2016/entries/qt-uncertainty/.

as to how the group might be affected. Ultimate outcomes may be positive: If a closed message conversation is in fact public, better understanding of these exchanges may help inform larger policy or technology.

At the same time, researchers also need to consider the ways the research project might affect the group participants or conversations negatively. For example, could the observation of researchers contribute to a "chilling effect" in terms of conversation? Will any group members have the sense they are performing for an external audience?

Does an increasing definition of closed messaging spaces as "public" also have repercussions? For example, does advertising or marketing within these spaces or with these member accounts now become acceptable? It also may be that in certain country contexts, where public exchanges have become more constrained due to legislation around speech, defining a closed group chat as "public" may have repercussions.

Lastly, among the group participants whose security researchers should consider are the researchers themselves: When entering into group conversations, advance planning can mitigate potential risks such as not using personal phone numbers for data collection.

Obligations: What are the Obligations for Researcher Disclosure and Informed Consent?

The chat space does not lend itself well to traditional practices of researcher disclosure and informed consent not only because of its encrypted nature. In addition, the ongoing inclusions of new members and departures of older ones create a challenge. Scholars Barbosa and Milan have argued, "[E]ven when one discloses their researcher identity, the fluidity of belonging and participation in a chat app creates a situation that it is not really possible to inform every group participant about the ongoing

observation, let alone remind each of it; ensuring 'informed' consent is probably just wishful thinking in very large groups."[86]

However, researcher disclosure and informed consent may not be required by different professional practices for all types of investigation. In the United States, for example, oral history, journalism, literary criticism, legal research, and historical scholarship, are not required to practice informed consent because they do not contribute to generalizable knowledge but instead specific knowledge of certain individuals.[87] Even within universities that have Institutional Review Boards, every academic research project does not need to be reviewed. The observation of public behavior, such as through election observation, can also fall under this exemption.[88] Practitioners of these fields may still self-regulate and define best practices and ethical codes that include informed consent.[89] In the European Union, in the wake of Nazi-era experimentation, free and informed consent has been emphasized in the context of medicine and biological research within the Nuremberg Codes as well as the European Charter of Fundamental Rights.[90]

Obligations: When Should Researchers Inform or Report Back to the Groups Involved?

The question of whether to report back is more important for Models 3 and 4, or projects that focus on more academic outcomes, to consider explicitly. Projects described in Models 1 and 2 for example, or based in journalism and civil society efforts such as election incident reporting, reported directly back to citizens as part of their design.

This question considers the traditions from which informed consent emerged. For example, take the case of the Tuskeegee

---

[86] Barbosa and Milan, "Do Not Harm in Private Chat Apps."

[87] Indeed, in this definition, the activities are not considered "research," which the U.S. Federal Register limits to activities of a systematic, generalizable nature. "Federal Policy for the Protection of Human Subjects," Federal Register (United States), Rules and Regulations, 82, no. 12 (January 19, 2017): 7149-7274.

[88] For one example, as defined in Emory University, "Emory University Institutional Review Board Policies and Procedures," September 18, 2020 update, http://www.irb.emory.edu/documents/PoliciesandProcedures.pdf.

[89] As for example, the Oral History Association, as referenced by American Historical Association, "Public Statement on Oral History and Human Subjects Regulation (November 2015)" (American Historical Association, November 12, 2015), https://www.historians.org/news-and-advocacy/aha-advocacy/public-statement-on-oral-history-and-human-subjects-regulation.

[90] "Charter of Fundamental Rights of the European Union."

Experiment in the United States, which demonstrated necessary limits within democracies on engaging in human subject research.

Beginning in 1932, the US Public Health Service conducted the "Tuskeegee Study of Untreated Syphilis in the Negro Male" for 40 years, which amounted to a passive observation of the effects of syphilis in hundreds of African-American men.[91] In doing so, the U.S. Public Health Service misled hundreds of participants and neglected to inform them of proper treatment options for the disease once their effectiveness became established, a clearly unethical research practice.

Especially when the topic of exchange is not merely misleading but potentially harmful — one of the motivations of this research outlined earlier — do researchers have an obligation to report their results back in certain cases? We know, for example, when events like data breaches occur, companies are now obligated to report them to those affected. At the same time, there may be thresholds to observe. Could reporting back to subjects, depending on the manner, have its own harmful consequences? Researchers must spend time trying to anticipate these scenarios, even if there are often unintended consequences difficult to foresee.

This question is not limited to research only in closed messaging applications, but in social media overall. The effort to establish reporting protocol, along with other practices such as anonymization, related to more public social media arenas still seems underway.[92]

---

[91] First captured October 11, 2007 at https://web.archive.org/web/20071011014512/http://www.cdc.gov/tuskegee/timeline.htm, "The U.S. Public Health Service Syphilis Study at Tuskeegee: The Tuskeegee Timeline," Centers for Disease Control and Prevention, http://www.cdc.gov/tuskegee/timeline.htm.

[92] As an example, Fiesler and Proferes, "'Participant Perceptions of Twitter Research Ethics"; for provocative considerations related to overall studies in Human-Computer Interaction, see Barry Brown et al., "Five Provocations for Ethical HCI Research," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery, 2016), 852–63, https://doi.org/10.1145/2858036.2858313.

Next Steps: Public Discussions amid an Evolving Technology

In the end, as researchers have noted as of 2020, it seems still the case that "using WhatsApp for research purposes (so far) is out of the ordinary"; as one research subject commented, "If I didn't know you [the researcher] in person, I would find it unusual" to use the interview research method in WhatsApp.[93] The interviewee also responded, "sending screenshots made me feel rather uncomfortable" and that "usually I would never let someone have such a deep look into my mobile phone."[94]

From descriptions of the research studies about closed messaging spaces starting to be available, we find evidence that many investigators are trying to follow ethical practices established by their various professional communities. They do this through practices that aim to protect the identities of chat participants. They also, for example, include considerations of possible current and future harm upon those observed. In recent years, expectations and requirements related to personal data collection and online storage have risen overall, making the obligation clearer for researchers across the board.[95]

The challenge, however, for closed messaging is that both the technology and its ethics are still emerging.[96] At bottom, we lack an understanding or approach to the issues posed by closed messaging that match the scale of its usage.

How severe is the problem of misinformation or disinformation within closed message chats, for example? Even the application owners do not know the answer to this question. Layers of encryption create hurdles that even internal engineers are unable to cross without trying to break the system. Due to technological

---

[93] Katja Kaufmann and Corinna Peil, "The Mobile Instant Messaging Interview (MIMI): Using WhatsApp to Enhance Self-Reporting and Explore Media Usage in Situ," Mobile Media & Communication 8, no. 2 (May 1, 2020): 229–46, https://doi.org/10.1177/2050157919852392.

[94] Kaufmann and Peil.

[95] Such as the Common Rule in the United States mentioned earlier or European Commission and Directorate General for Research, Ethics for Researchers: Facilitating Research Excellence in FP7. (Luxembourg: Publications Office, 2013), http://bookshop.europa.eu/uri?target=EUB:NOTICE:KI3213114:EN:HTML.

[96] See for example a discussion on citizen journalism understanding of ethics in Shepherd Mpofu and Shanade Bianca Barnabas, "Citizen Journalism and Moral Panics: A Consideration of Ethics in the 2015 South African Xenophobic Attacks," African Journalism Studies 37, no. 4 (November 2016): 115–36, https://doi.org/10.1080/23743670.2016.1256053.

barriers and ethical hesitance, current studies of messages flowing through these spaces have been very small, which means they are necessarily limited in terms of their ability to speak to larger trends. Yet what research exists is troubling in terms of misinformation's impact within closed applications.[97]

A central conundrum is that the technology allows for scaled social interaction, far beyond the intimacy that encrypted chat suggests. In order to begin to answer the academic question of larger impact, Models 3 and 4 are attempting more generalizable conclusions about the "public."[98] A further complication to the research is the potential variance in country or cultural expectations across *multiple publics,* let alone from different group participants, in terms of research taking place within their chats. Perhaps, in other words, the expectations for privacy or the public-private divide is different according to topic, culture, and country.

Given that closed messaging technology and research continues to evolve, there are also a series of needed discussions suggested by this landscape review: with the public (or publics), with companies, and with regulators.

> **Discussions with the public.** When closed messaging research is in the public interest, it seems worth asking whether or not key questions might be asked directly to representatives of the communities or countries involved. For example, what ways might establish better understandings of community, cultural, or country expectations around public versus private conversation? In order to be representative, methods such as surveys might be employed. Efforts to gain and assess public feedback may be difficult. However, they also offer an opportunity for collaborating across different organizations or disciplines in order to achieve some consensus related to research in closed messaging spaces.

---

[97] Josemar Alves Caetano et al., "Characterizing Attention Cascades in WhatsApp Groups," in *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19 (New York, NY, USA: Association for Computing Machinery, 2019), 27–36, https://doi.org/10.1145/3292522.3326018.

[98] A goal as outlined in Halford et al., "Understanding the Production and Circulation of Social Media Data."

Discussions with companies. A key issue when it comes to any social media research is that most approaches do not have direct access to app data (partnership with corporate research efforts offer some exceptions). In the case of closed messaging applications, the potential for that exchange between companies that own these applications and researchers is inherently limited; encryption makes this so. Furthermore, self-reported or company-provided data requires a different methodological approach compared to externally audited information. Perhaps, for example, greater clarity about the boundaries between private and public space could be jointly discussed and communicated.

Discussions with regulators. Given the ways that these spaces are by design both public and private — at times one more than the other — an improved exchange among publics, companies, and researchers may lead to creative approaches and answers to all the questions outlined above. They also need to include the specific perspectives of industry regulators. Defining with more specificity what the meaning of what a "good faith" research effort is in collaboration with industry, per the Code of Practice on Disinformation, could be a start.[99] Collaboration would be especially helpful with regard to boundaries of advertising in closed messaging spaces as well.[100]

At bottom, we are confronted with the challenge not only of an evolving technology, but an evolving concept of publics and public conversation. Legal scholarship reminds us that the "public domain," for example, is "a metaphor, a social-legal construct"; it has:

---

[99] European Union, "European Union Code of Practice on Disinformation"; Florian Saurwein and Charlotte Spencer-Smith, "Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe," *Digital Journalism* 8, no. 6 (July 2, 2020): 820–41, https://doi.org/10.1080/21670811.2020.1765401; James Pamment, "EU Code of Practice on Disinformation: Briefing Note for the New European Commission," Partnership for Countering Influence Operations, Policy Perspectives Series (Carnegie Endowment for International Peace, March 2020), https://carnegieendowment.org/2020/03/03/eu-code-of-practice-on-disinformation-briefing-note-for-new-european-commission-pub-81187.

[100] Recent concerns about WhatsApp's upcoming Privacy Policy change, and particularly its efforts to support business exchanges in the application, have resulted in a delay to its original deadline of February 8, 2021. WhatsApp, "Giving More Time for Our Recent Update," January 15, 2021, https://blog.whatsapp.com/giving-more-time-for-our-recent-update/?lang=en.

"an instrumental purpose — to assist us in thinking of a complex issue, to organize our thoughts, to serve as a 'short cut' to denote a mindset, a view, a perception" about the legal status of different types of information and what can be done with this information."[101]

The approaches to research on closed messaging apps draw from our understanding of publics. Simultaneously, these research studies will also shape our understanding of publics and public discussion.

An increase in adoption of closed messaging apps will be accompanied by an increase in research studies around them. Each research study will negotiate the balance between public and private treatment of closed messaging spaces differently, but there is a need to converge on foundational principles that can guide these research decisions. This paper lays out an opening framework to develop these principles and is an invitation to further dialogue and deliberation toward research norms, norms that protect our democracies in action as well as the people who provide insights into them.

[101] First quote, Pamela Samuelson, "Enriching Discourse on Public Domains," Duke Law Journal 55, no. 4 (February 1, 2006): 783–834; second quote, Michael Birnhack in Samuelson.

Appendix A: Closed Messaging Data Collection and Sharing Practices

The paper goes over the motivation for research on closed messaging spaces and the ethical conundrums when conducting such research. This paper has crystallized six questions around research norms related to closed messaging apps that need broader deliberation. The principles guiding these deliberations, however, are tied to specific operational choices when it comes to practice.

In this section, we detail the design choices a researcher or developer will need to make when extracting data from closed messaging apps. Closed messaging apps require that the researchers sign up with a specific phone number. Depending on the platform studied, and the research set up, relevant data could be collected through a mobile phone or a web client. In this section, we describe different design decisions in operationalizing research on closed messaging apps, and their interaction with the six questions mentioned in the paper. We also present three possible configurations of research methodology, with each balancing researcher interests with chat app user safety differently.

Once decisions have been made, one overall practical suggestion for researchers is to submit their research protocols to an external trusted colleague for review. Whether or not researchers submit their protocols to Independent Review Boards, projects involving research in the public interest can benefit from peer feedback related to security and obligations.

The design choices in setting up the data collection process have five main stages:

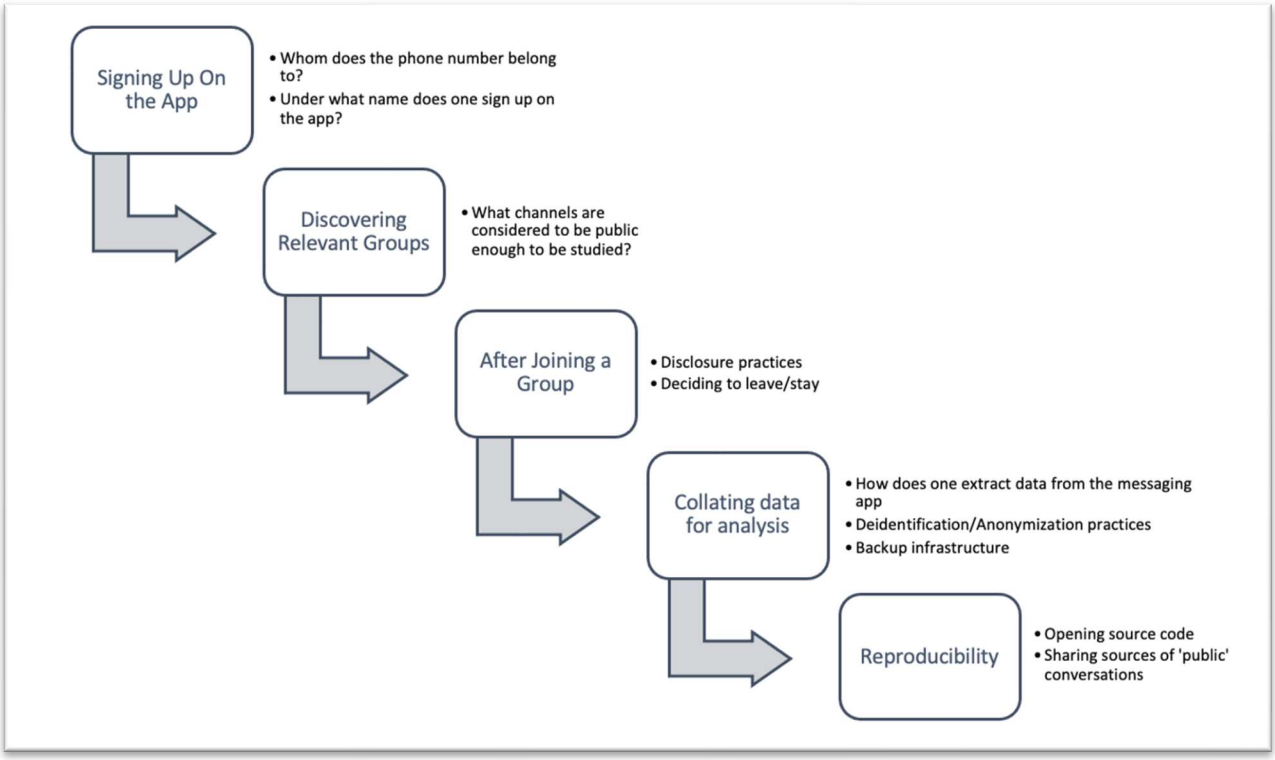
Signing Up On the App
• Whom does the phone number belong to?
• Under what name does one sign up on the app?
Discovering Relevant Groups
• What channels are considered to be public enough to be studied?
After Joining a Group
• Disclosure practices
• Deciding to leave/stay
Collating data for analysis
• How does one extract data from the messaging app
• Deidentification/Anonymization practices
• Backup infrastructure
Reproducibility
• Opening source code
• Sharing sources of 'public' conversations

Figure 1. Five stages of design choice.

Signing Up on The App

Relevant Questions

- Security: How might the research project affect group members and their conversations?
- Obligations: What are the obligations for researcher disclosure and informed consent?

In order to collect data from a messaging app, a researcher or developer must create an account using a phone number. At this stage, the individual/team must decide on:

- **What phone number to use to register**

  The phone number could be procured by researchers as individuals or by an organization. Phone numbers can be seen by everyone in a group conversation and might be used for

digital marketing and other unsolicited communication. Separating phone numbers used for research purposes and for personal communication can add the necessary separation between researchers' personal lives and the research work.

- **How to name the account**

  At the time of signing up on a messaging platform, a user has to declare a name that identifies them to others on the app. Researchers may choose to use individual names or the name of the research group/study. Declaring the name as a research study or organization might make it easier for group admins and participants to identify the researcher, enabling more informed and meaningful consent in the admission of researchers to the group.

  Researchers may provide additional description of their purpose in the 'Status' or 'About' field, if the platform allows it. Using language that research subjects understand can result in more meaningful consent. This, however, may be difficult in multilingual groups. Limit on the length of the text description restricts lengthy or multilingual description.

Regardless of the username declared on the messaging app, note that the owner of a phone number may still be identified through apps such as True Caller. For researchers working in sensitive contexts, identification may be another consideration to have phone numbers procured through an organization.

Discovering Relevant Groups

Relevant Questions

- Parameters: Exactly when is a closed message chat "public"?
- Security: How might the research project affect group members and their conversations?

The paper brings to the fore a conundrum between public and private groups on messaging apps. This conundrum is best reflected in the process of discovering group conversations.

On apps such as Telegram, "public channels" can be found through searching for the channel names on the app. WhatsApp, however, does not provide this option. Links to group conversations must be found by proactive searching for group links online. This search could span Facebook groups, Twitter conversations or a broad search on the world wide web. While Telegram aggregates popular channels on the app[102], several websites (not endorsed by

[102] "Telegram Channels List: Discover Interesting Channels for Your Telegram," tlgrm.ru, accessed January 22, 2021, https://tlgrm.eu/channels.

WhatsApp) aggregate WhatsApp group links.[103] Links to new group chats may also be shared within a group that researchers are monitoring.

The "publicness" of a group conversation is assumed from the group administrators' intent to make it discoverable. While on Telegram this intent is explicit in creation of a channel, on WhatsApp and Signal it is less so. The latter two platforms provide an option to invite other members to the group conversation via a link. This feature is used by individuals to invite people to a group intended for private communication, as well as by those intending to create a public group. The two purposes cannot be easily distinguished from the group link alone.

The nature of the online forum on which the message group link is found can provide some indication of the intention with which these groups were created. Message groups shared exclusively on closed Facebook groups, for example, should be assumed to be directed at members of a closed community. Group links shared on public Twitter profiles are plausibly created for broader engagement.

With group links aggregated by independent websites, the intended nature (public versus private) of the group is hard to discern. The theme under which the group is categorized on these sites could be used to filter groups for research. For example, groups categorized as "regional news" or "shopping deals" are possibly intended as broadcast groups, while groups tagged as "study groups" might be intended for a more selective audience. The categorization of groups on a website is decided by website creators and not the message group admins or participants. Thus, researchers must use their discretion in using these websites for research purposes.

Finally, researchers may consider joining additional groups through links shared on groups they are monitoring. The public status of these groups is derived from the perceived status ascribed to the message group they were found on. If the assumption is that the public status of all messaging groups is unclear to begin with, the public status of groups found on other messaging groups is doubly weakened.

[103] Du B, "10000+ Active WhatsApp Group Links | December 2020 [Updated]," WhatsApp Groups (blog), December 19, 2020, https://whatsappgroups.app/whatsapp-group-links/.

After Joining A Group Conversation

Relevant Questions

- Parameters: Exactly when is a closed message chat "public"?
- Security: How might the research project affect group members and their conversations?
- Obligations: What are the obligations for researcher disclosure and informed consent?

Disclosure Norms

On joining a conversation, researchers may choose to disclose their purpose. Researchers at COMPROP declared their presence in groups at the time of joining and were often removed from the groups.[104] Researchers at MIT IDSS and Columbia University's Tow Center did not declare their purpose for joining the group.[105]

Deciding to Leave a Group

After joining a group, researchers may realize that the group is intended to be private. Researchers at The Tow Center used group size as a proxy for private or public nature of the group and only studied groups that had more than 60 members. The number of people in a message group can be found only after joining it. Since the size of a group dynamically changes, researchers may choose to keep a time window after joining, before deciding to study the group or leave it. Thoughtful protocols may be followed in this time window to respect privacy and consent of group participants.

[104] Vidya Narayanan et al., "News and Information over Facebook and WhatsApp during the Indian Election Campaign."

[105] Garimella and Tyson, "WhatsApp, Doc?"; Bengani, "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It."

Extracting and Collating Data for Analysis

Relevant Questions

- Parameters: Who does the data belong to?
- Security: What is the plan for storing closed messaging research data in a secure fashion?

Method of Data Collection

After joining group conversations, researchers need to extract the content from the messaging app via a mobile app or a web client for structured analysis. On encrypted messaging services such as WhatsApp and Signal, media items can only be extracted once they are downloaded to a device. Thus, researchers must go through all media items and load them so that they are downloaded on a mobile device or a computer.

To collect data stored on phones, researchers may use a variety of methods, each of which results in different granularity and quality of data. The method adopted would also depend on the functionalities provided by the messaging platforms. The methods could be:

- Rooting phones to decrypt local app database:

  Encrypted messaging apps store all the conversations and additional information needed for app operation on the phone in an encrypted database. While this database can be easily extracted, the key to decrypt this database can be hard to retrieve. One way to extract this key is to gain "root access" to the phone.

  This method was used by the MIT IDSS group for extracting content from WhatsApp. This method results in the most granular data on message conversations. It allows researchers to access unique identification of every sender, the membership list in different groups. In the case of WhatsApp, one may even have access to the media ID and hash generated by WhatsApp. Researchers might also access fields not visible on the mobile app but created by the platform for network or content analysis such as whether a post was spam or not, or the number of times it has been forwarded. To extract the data, the local phone database must be routinely copied to a computer on which it can be decrypted.

- Backing up the phone

  Messaging platforms may provide users an option to back up their conversations. These backups may be encrypted by a key available to the user (as on Signal). In some cases, the backups may not be encrypted. In older versions of WhatsApp, backup on iPhone was not encrypted.

- Manually copying content to a database

  This is a labor-intensive process where researchers copy every message or media item to a database. This database could be a spreadsheet hosted locally or on a cloud service provider. This method may not be feasible for long-term or high-volume data collection.

- Exporting Chat

  Depending on the jurisdiction, closed messaging platforms may provide users an option to export a plain text or html file of their chat history. The chat can be exported and shared over email, a cloud storage provider, or another mobile app.

  On WhatsApp, exporting a chat results in a text file with links to media items. The exported text file results in the timestamp of the message, the sender (as they have identified themselves on the app) and the message itself. The exported file may also unexpectedly skip many of the media items. While this method is less laborious than the previously mentioned approaches, it relies on the platforms' export chat feature that may not be robust. WhatsApp places a restriction on the number of media items that can be exported through the export chat feature.

- Researchers may also extract content from web clients of messaging apps. Web-based methods include reverse-engineering the web application API or crawling from the user interface.

De-identification and Anonymization

Public or not, message groups are a space for consequential discourse. People may express political views or share their health histories. All the message conversations are tied to a phone number that could be used to link these conversations to other databases. Truly anonymous data collection may not be possible; researchers may consider never saving raw data it parses, but this may still leave traces.[106] It is imperative that at the very least the data is de-identified.

Deidentification does not guarantee anonymization, but as more data fields are de-identified, the risk of deanonymization decreases. In addition to phone numbers, researchers may also consider anonymizing the group conversation name. If the focus is only the message content, researchers may choose to use a different anonymization "seed" for new batch of data. This would delink conversations from the same sender across different data snapshots.

---

[106] On the difficulties of anonymization, a classic starting place is Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," UCLA Law Review 57, no. 6 (August 2010), https://www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/

Future research could test the technical limits of anonymization of data from closed messaging apps.

Backup

Once the data has been anonymized, researchers must decide on how, if at all, to preserve the original de-anonymized data. Depending on the volume of data, it may be possible to save the data on local external drives. With high volumes of data, researchers could spin up their own servers or alternatively rely on cloud service providers. With cloud service, researchers are relying on security features of their service providers.

Reproducibility

Reproducibility is an important expectation in research. As this paper has contended, in some circumstances closed messaging apps may be assumed to be sufficiently public to merit research. The groups on these platforms, however, are not public in the sense that all data can be shared under open access (as can be reasonably done for Reddit or Twitter conversations). Researchers may consider opening their data through restricted access.

The method adopted for extracting data, anonymizing it, and analyzing data should be clarified in publications using the data. Researchers may also consider opening the source code used to extract and analyze the data. Documenting ethical practices of data collection in the documentation of the code could remind others who use the code to take adequate measures to protect the subjects they are studying.

Scenarios

We have explained how design decisions at different stages of the data collection and preparation process may balance research goals with protection of group participants. At present, there is no single agreed upon sequence of decisions to strike the appropriate balance. The appropriate balance depends on the research setting.

These design choices can be seen as a series of levers whose joint configuration results in a specific balance of researcher interest and app user safety. To help clarify the implications of design choices, we present three possible configurations.

Scenario One: Minimizing Data Collection

Step	Approach
Signing Up on the App	Register number to research organization
Discovering Relevant Groups	Only group links shared on public accounts
After Joining Group	Announce research purpose
Collating Data for Analysis	Manually copy only media items from individual message conversations into a separate Excel sheet
Reproducibility	Share links and media items externally

This approach minimizes data collection at every step in the interest of honoring app users consent to be researched. It assumes that only group links broadcast on a public profile are public enough to be researched. In this scenario, user privacy is prioritized and no phone numbers are collected or stored in a database. All the data is saved locally to minimize data leaks from breaches of cloud service providers.

Scenario Two: Journalism Investigation

Step	Approach
Signing Up on the App	Register to individual researcher
Discovering Relevant Groups	Shared on public accounts, as well as sent by individuals
After Joining Group	Do not declare presence
Collating Data for Analysis	Extract from backup of message conversations but never put data on cloud service providers
Reproducibility	Don't share data externally

This is a possible approach taken by a journalist pursuing investigative work for a limited duration.

In this scenario, access to conversations is prioritized over disclosures. While the number is registered to an individual journalist, they don't declare their presence to minimize their effect on group dynamics and prevent harms from self-identification. All content collected in this process is treated as sensitive. In the interest of protecting sources (group conversation participants), the journalist would not put the research data on the cloud or share the data externally.

Scenario Three: Quantitative Project

Step	Approach
Signing Up on the App	Register to individual organization
Discovering Relevant Groups	Shared on public accounts, as well as group sites
After Joining Group	Do not declare presence
Collating Data for Analysis	Extract from backup of message conversations as well as export chat
Reproducibility	Restricted access to data. All code is open

This approach may be taken by a quantitative researcher seeking to do long-term data analysis on group conversations. This is possibly a large-scale research project. The scale of data collection merits investment in a technical infrastructure that ensures security and strong anonymization of data.

Discussion

In each of the scenarios described, researchers define some conditions under which the closed message groups are considered public enough to be studied. As the definition of public is expanded to include more groups or increase the granularity of data collection, the burden to ensure security of the data also increases. Researchers must gauge their ability to securely analyze and share the volume of data they collect in their research.

Research norms on closed messaging apps are still evolving; the criteria discussed in this Appendix will be continually updated at: https://tattle.co.in/products/closed-messaging-checklist/.

However, by codifying responsible data management practices prior to the study, researchers can anticipate and mitigate risks to their research subjects. In absence of an Institutional Review Board, researchers could have their research protocol reviewed by professional peers, not tied to the research project. Researchers should also consider consulting technical experts for secure data collection and storage practices.

Appendix B: Examples of Considerations, Practices, and Questions Regarding Closed Message Research

Past convenings and guides include a wealth of example questions and considerations that can help researchers think more reflectively and carefully about their proposed projects into social media spaces overall, and in closed messaging applications in particular. Include below are quotes and questions within these papers, as well as distilled insights from them.

Considerations from Social Media Data Stewardship, 2016[107]

- Is informed consent necessary/recommended when collecting social media data? (Consider public vs. private data)
- How can informed consent be collected [when collecting social media data through an API]?
- When is informed consent not necessary?
- Should academic researchers be held to a higher [ethical] standard than industry or government?
- Where and how is data collected?
- Time restrictions on data availability?
- What is an appropriate sampling/recruitment strategy to ensure an accurate representation of social media users and their views?
- What are the issues associated with publishing or sharing social media datasets that you collected for your research?
- What metadata schemas and what formats should be used to preserve social media data?
- What are the considerations for different data types such as posts, profiles, interactions, social networks [as platforms evolve]?
- What are the main strategies to ensure data reuse? For example, should researchers preserve all available metadata ("just in case"/"all you can get" approach)?

[107] Anatoliy Gruzd, Jenna Jacobsen, and Elisabeth Dubois, "Challenges and Opportunities of Doing Research with Social Media Data," Social Media Data Stewardship (blog), accessed December 3, 2020, https://socialmediadata.org/2016/04/16/challenges-and-opportunities-of-doing-research-with-social-media-data/.

Considerations from Frankel, 2018[108]

"News organizations' audiences are increasingly moving from public social media to closed or semi-closed platforms like WhatsApp, Discord, and Facebook Groups. But there are still opportunities for good reporting on the communities we cover."

- How easy and appropriate is it to join groups and communities intent on peer-to-peer online discussion?
- What issues or barriers are there in journalists building meaningful connections in these online spaces?
- How much of what can be observed in these forums would a journalist wish to use or relay to colleagues for publication or broadcast?
- "Proceed with caution: False information in WhatsApp groups is ubiquitous and has horrific real-life consequences."
- What do you know about the community you're seeking to find? Do you have an invitation to a server already? Are you looking for a story on a whim or have you gotten a tip from a Redditor, for example?
- Are you able and willing to run your account transparently as a journalist and link to your wider digital profile?
- If so, how will you protect yourself against possible recriminations when you find someone you'd like to approach? And, if not, how can you justify your anonymity at the outset of your search and what will you do when you want to quote or reach out to a user?

Considerations from Sehat, 2018[109]

"We wanted to inquire about two things, which exist in apparent tension: What are the good methodologies to explore this space — in which one could expect reliable data? — and what ethics of the situation need to be taken into account, given the very privacy that grounds these spaces?"

- We need to gather more information about what sort of content gets shared and who shares it.
- Before deploying surveys or other research, think about the limits of what knowledge could be reliably gathered.
- The bar for ethical review may need to be higher than average, given the context.

---

[108] Mark Frankel, "The Promises and Pitfalls of Reporting within Chat Apps and Other Semi-Open Platforms: A Journalist's Guide," Nieman Lab, July 10, 2018, https://www.niemanlab.org/2018/07/a-journalists-guide-to-the-promises-and-pitfalls-of-reporting-within-open-and-closed-and-semi-open-platforms/.

[109] Sehat, "Closed Messaging Spaces."

Considerations and Practices from Bengani, 2019[110]

- Don't use your personal phone or phone number.
- Identify yourself and your institutional affiliation.
- Use hard drives and not a cloud in data collection.
- Respect general privacy and have ethical cautions.

Questions from Dotto et al., 2019[111]

- Are there ways of obtaining the information you are seeking without entering into closed online spaces?
- What do you hope to obtain from joining this group? Are you looking to find sources and tips, or to gain background knowledge to inform your reporting? Or are the existence and the content of the group itself the focus of your intended story?
- Is this a group that would expect "lurkers"? Would members reasonably expect conversations and other content from these groups to be made public?
- What is the size of each closed group you are planning on entering, and how does that affect the expectation of privacy for each group?
- Would your writing expose group members to negative consequences?
- What is the public interest in your potential story?
- Are you planning on entering multiple groups? What is the minimum number of closed spaces you can enter into to find the information you need?
- Will you disclose your identity? Will you disclose when you first enter the group, when you find something useful in the group you'd like to include in your reporting, when you have completed newsgathering in the group, or when your story is published? If you plan on being in the group for an extended period, or the group gains new members after your initial disclosure, will you re-disclose your identity?
- What is the purpose of this group? Is the group likely to be hostile, and how would group members react to a reporter within their midst? Entering a closed group that facilitates criminal activity or advocates extremist ideologies, for example, may lead to a different disclosure decision than entering a WhatsApp conversation consisting of local parents or a secret group of employees looking to unionize.
- Is your entry and presence in the group, using your real identity, likely to draw unwanted attention or abuse? Journalists of color and women, for example, may face additional

[110] Bengani, "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It."

[111] Carlotta Dotto, Rory Smith, and Claire Wardle, "Closed Groups, Messaging Apps, and Online Ads," First Draft's Essential Guide (First Draft, November 2019), https://firstdraftnews.org/wp-content/uploads/2019/11/Messaging_Apps_Digital_AW-1.pdf?x80491.

security concerns when entering into certain potentially hostile groups, which may lead to a different disclosure decision.

- If you decide to enter the group using your real identity, to whom will you disclose this information? Will you disclose it to the group administrator, or to the whole group?
- Will you also disclose your reasons for being in this group?
- If the group requires you to answer certain questions before admission, will you answer these questions honestly?
- If there are explicit confidentiality clauses in the community guidelines of the groups you are entering: How you are going to describe the methods of newsgathering in the resulting story?
- Will you go back into the group after the story's publication and share the information you have learned?

Considerations from Gibson, 2020[112]

"There would seem to be obvious value in exploring the use of digital (ethnographic) data collection methods in youth research. However, it is important to approach this novel methodological terrain from a reflexive standpoint that can illuminate how this technology shapes the research encounter and the nature of the data collected."

- What are ethical boundaries to consider when leaving open instant messaging chats and conversations before and after the end of an interview?
    - Can this boundary or restriction lead to lack of insight regarding qualitative research in closed messaging spaces?
    - Is it the participant's or researcher's responsibility or decision to keep the conversation open or to be archived?
    - The additional risk of intrusions into the researcher's privacy through the extended access the participants had to engage were considered insignificant given that this was a phone used and monitored exclusively for research purposes.
- How will manage the participant's potential distress and risks associated with collecting qualitative data and quantitative data in such spaces?

Considerations from Kaufmann and Peil, 2020[113]

"Although offering a range of opportunities for qualitative research—in situ exploration; contextual orientation; insights into people's experiences, perceptions, and feelings, to name but a few—both diary and mobile experience sampling methods have some limitations with regard to

[112] Kerry Gibson, "Bridging the Digital Divide: Reflections on Using WhatsApp Instant Messenger Interviews in Youth Research," Qualitative Research in Psychology, April 17, 2020, 1–21, https://doi.org/10.1080/14780887.2020.1751902.

[113] Kaufmann and Peil, "The Mobile Instant Messaging Interview (MIMI)."

the depth of exploration and range of communication possibilities between researcher and respondent in the situations surveyed."

- The implementation of MIMIs, also into other fields of research, needs first and foremost the target group to be taken into account carefully.
- Contact potential participants directly on WhatsApp and obtain their informed consent to make use of their data for academic purposes and, on the part of the research, assuring their anonymity and the data's careful handling.
- Be clear that you were neither aware of, nor had any influence on, the purposes and objectives that the platform operator was pursuing with customer data.
- Using a communication space perceived as private might be advantageous for a rather unobtrusive topic like everyday media use. More serious or intimate topics might necessitate either other measures to gain trust (see also participants perspective in what follows) or, contrastingly, call for establishing a professional distance.
    - How and to what extent will you protect the participants' privacy when interviewing?
    - What environmental and external factors will you take into consideration that will ensure the participants private outside of the closed messaging platform?
- To what extent should academic research use commercial software based on private interests such as closed messaging spaces being used in research?
- Should researchers exclude topics that are likely to produce sensitive data when using a commercial service? Or is it quite the contrary, that appropriating the informal and casual atmosphere of a private communication space, like WhatsApp, allows researchers to gain data that users would not disclose in a more formal research setting?

Appendix C: References and Additional Resources

American Historical Association. "Public Statement on Oral History and Human Subjects Regulation (November 2015)." American Historical Association, November 12, 2015. https://www.historians.org/news-and-advocacy/aha-advocacy/public-statement-on-oral-history-and-human-subjects-regulation.

Anderson, Meghan Keaney. "Why We're Thinking About Messaging Apps All Wrong." *HubSpot* (blog), February 9, 2016. https://blog.hubspot.com/marketing/messaging-apps.

Arun, Chinmayi. "Opinion | India May Be Witnessing the next 'WhatsApp Election' — and the Stakes Couldn't Be Higher." *Washington Post*, April 25, 2019. https://www.washingtonpost.com/opinions/2019/04/25/india-could-see-next-whatsapp-election-stakes-couldnt-be-higher/.

B, Du. "10000+ Active WhatsApp Group Links | December 2020 [Updated]." WhatsApp Groups (blog), December 19, 2020. https://whatsappgroups.app/whatsapp-group-links/

Barbosa, Sérgio, and Stefania Milan. "Do Not Harm in Private Chat Apps: Ethical Issues for Research on and with WhatsApp." *Westminster Papers in Communication and Culture* 14, no. 1 (August 14, 2019): 49–65. https://doi.org/10.16997/wpcc.313.

Baumgartner, Jason, Savvas Zannettou, Megan Squire, and Jeremy Blackburn. "The Pushshift Telegram Dataset." *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 26, 2020): 840–47.

Bengani, Priyanjana. Letter to Aleksei Kaminski. "Email Reply: Introducing Myself and Your Research," March 18, 2020.

———. "India Had Its First 'WhatsApp Election.' We Have a Million Messages from It." *Columbia Journalism Review*, October 16, 2019. https://www.cjr.org/tow_center/india-whatsapp-analysis-election-security.php.

Biswas, Soutik. "Fighting India's WhatsApp Fake News War." *BBC News*, August 20, 2018, sec. India. https://www.bbc.com/news/world-asia-india-45140158.

Brown, Barry, Alexandra Weilenmann, Donald McMillan, and Airi Lampinen. "Five Provocations for Ethical HCI Research." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 852–63. New York, NY, USA: Association for Computing Machinery, 2016. https://doi.org/10.1145/2858036.2858313.

Caetano, Josemar Alves, Gabriel Magno, Marcos Gonçalves, Jussara Almeida, Humberto T. Marques-Neto, and Virgílio Almeida. "Characterizing Attention Cascades in WhatsApp Groups." In *Proceedings of the 10th ACM Conference on Web Science*, 27–36. WebSci '19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3292522.3326018.

Caetano, Josemar Alves, Jaqueline Faria de Oliveira, Helder Seixas Lima, Humberto T. Marques-Neto, Gabriel Magno, Wagner Meira Jr, and Virgílio A. F. Almeida. "Analyzing and Characterizing Political Discussions in WhatsApp Public Groups." *ArXiv:1804.00397 [Cs]*, April 2, 2018. http://arxiv.org/abs/1804.00397.

Centers for Disease Control and Prevention. "The U.S. Public Health Service Syphilis Study at Tuskeegee: The Tuskeegee Timeline." Accessed July 6, 2020. http://www.cdc.gov/tuskegee/timeline.htm.

"Charter of Fundamental Rights of the European Union." *Official Journal of the European Union*, no. C 83 (March 30, 2010): 389–403.

Cheeseman, Nic, Jonathan Fisher, Idayat Hassan, and Jamie Hitchen. "Social Media Disruption: Nigeria's WhatsApp Politics." *Journal of Democracy* 31, no. 3 (July 14, 2020): 145–59. https://doi.org/10.1353/jod.2020.0037.

Cox, Joseph. "VICE - Google Is Letting People Find Invites to Some Private WhatsApp Groups." *Motherboard (Vice)*, February 21, 2020. https://www.vice.com/en_us/article/k7enqn/google-is-letting-people-find-invites-to-some-private-whatsapp-groups.

"Digital Rights in Africa Report 2019." *Paradigm Initiative*. Accessed December 3, 2020. https://paradigmhq.org/download/dra19/.

Dotto, Carlotta, Rory Smith, and Claire Wardle. "Closed Groups, Messaging Apps, and Online Ads." First Draft, November 2019. https://firstdraftnews.org/wp-content/uploads/2019/11/Messaging_Apps_Digital_AW-1.pdf?x80491.

Dwyer, Maggie, Jamie Hitchen, and Thomas Molony. "Between Excitement and Scepticism: The Role of WhatsApp in Sierra Leone 2018 Elections." In *Social Media and Politics in Africa: Democracy, Censorship and Security*, edited by Maggie Dwyer and Thomas Molony. Zed Books Ltd., 2019.

Emory University. "Emory University Institutional Review Board Policies and Procedures," September 18, 2020. http://www.irb.emory.edu/documents/PoliciesandProcedures.pdf.

European Commission and Directorate General for Research. *Ethics for Researchers: Facilitating Research Excellence in FP7.* Luxembourg: Publications Office, 2013. http://bookshop.europa.eu/uri?target=EUB:NOTICE:KI3213114:EN:HTML.

European Union. "European Union Code of Practice on Disinformation." European Union, September 26, 2018.

———. General Data Protection Regulation (GDPR), REGULATION (EU) 2016/ 679 § Art. 4 GDPR - Definitions (2018). https://gdpr.eu/article-4-definitions/.

Fabricio Benevenuto. Letter to Aleksei Kaminski and Connie Moon Sehat. "Introducing Myself and Your Research- Aleksei," September 3, 2020.

"Federal Policy for the Protection of Human Subjects." *Federal Register (United States)*, Rules and Regulations, 82, no. 12 (January 19, 2017): 7149-.

Fiesler, Casey, and Nicholas Proferes. "'Participant' Perceptions of Twitter Research Ethics." *Social Media + Society* 4, no. 1 (January 1, 2018): 2056305118763366. https://doi.org/10.1177/2056305118763366.

Frahm-Arp, Maria. "The Political Rhetoric in Sermons and Select Social Media in Three Pentecostal Charismatic Evangelical Churches Leading up to the 2014 South African Election." *Journal for the Study of Religion* 28, no. 1 (2015): 115–41.

Frankel, Mark. "The Promises and Pitfalls of Reporting within Chat Apps and Other Semi-Open Platforms: A Journalist's Guide." *Nieman Lab*, July 10, 2018. https://www.niemanlab.org/2018/07/a-journalists-guide-to-the-promises-and-pitfalls-of-reporting-within-open-and-closed-and-semi-open-platforms/.

Freetown City Council. "FCC to Start Digital Town Hall Initiative — Freetown City Council." *Freetown City Council* (blog), September 24, 2020. https://web.archive.org/web/20201021133638/https://fcc.gov.sl/fcc-to-start-digital-town-hall-initiative/.

Garimella, Kiran. Letter to Connie Moon Sehat. "Re: [Air-L] [Request for Input] Research Practices for Closed Messaging Groups," March 25, 2020.

Garimella, Kiran, and Gareth Tyson. "WhatsApp, Doc? A First Look at WhatsApp Public Group Data." In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*. Stanford, CA: The AAAI Press, 2018. https://aaai.org/Library/ICWSM/icwsm18contents.php.

Garrison, Philip. Letter to Connie Moon Sehat. "[Request for Clarification] Research Practices for Closed Messaging Groups," July 22, 2020.

Ghoshal, Abhimanyu. "WhatsApp Launches a Tip Line in India to Battle Fake News." *The Next Web*, April 2, 2019. https://thenextweb.com/apps/2019/04/02/whatsapp-launches-a-tip-line-in-india-to-battle-fake-news-ahead-of-national-elections/.

Gibson, Kerry. "Bridging the Digital Divide: Reflections on Using WhatsApp Instant Messenger Interviews in Youth Research." *Qualitative Research in Psychology*, April 17, 2020, 1–21. https://doi.org/10.1080/14780887.2020.1751902.

Gil de Zúñiga, Homero, Alberto Ardèvol-Abreu, and Andreu Casero-Ripollés. "WhatsApp Political Discussion, Conventional Participation and Activism: Exploring Direct, Indirect and Generational Effects." *Information Communication and Society*, July 20, 2019, 1–18. https://doi.org/10.1080/1369118X.2019.1642933.

Goel, Vindu. "In India, Facebook's WhatsApp Plays Central Role in Elections." *The New York Times*, May 16, 2018, sec. Technology. https://www.nytimes.com/2018/05/14/technology/whatsapp-india-elections.html.

Gruzd, Anatoliy, Jenna Jacobsen, and Elisabeth Dubois. "Challenges and Opportunities of Doing Research with Social Media Data." *Social Media Data Stewardship* (blog). Accessed December 3, 2020. https://socialmediadata.org/2016/04/16/challenges-and-opportunities-of-doing-research-with-social-media-data/.

Guess, Andrew M., and Benjamin A. Lyons. "Misinformation, Disinformation, and Online Propaganda." In *Social Media and Democracy: The State of the Field, Prospects for Reform*, edited by Joshua A. Tucker and Nathaniel Persily, 10–33. SSRC Anxieties of Democracy. Cambridge: Cambridge University Press, 2020. https://www.cambridge.org/core/books/social-media-and-democracy/misinformation-disinformation-and-online-propaganda/D14406A631AA181839ED8969165985oo.

Haleva-Amir, Sharon. "Not All about That Facebook: Political Campaigns and Civic Engagement in the 2015 Elections." *Israel Affairs* 22, no. 3/4 (July 2016): 711–26. https://doi.org/10.1080/13537121.2016.1174375.

Halford, Susan, Mark Weal, Ramine Tinati, Les Carr, and Catherine Pope. "Understanding the Production and Circulation of Social Media Data: Towards Methodological Principles and Praxis." *New Media & Society* 20, no. 9 (September 2018): 3341–58. https://doi.org/10.1177/1461444817748953.

Hassan, Idayat, and Jamie Hitchen. "Forums of Debate? WhatsApp and the Gambia's Political Transition." Centre for Democracy and Development, n.d. https://cddelibrary.org/wp-content/uploads/2020/09/FORUMS-OF-DEBATE-WHATSAPP-AND-THE-GAMBIAS-POLITICAL-TRANSITION.pdf (now at https://docplayer.net/195505832-Forums-of-debate-whatsapp-and-the-gambia-s-political-transition-by-idayat-hassan-and-jamie-hitchen-september-2020.html).

Hegde, Pranav. "Coronavirus Impact | WhatsApp Global Usage Spikes by 51% during Pandemic." *MSN News: Money Control*, March 26, 2020. https://www.msn.com/en-in/news/other/coronavirus-impact-whatsapp-global-usage-spikes-by-51-25-during-pandemic/ar-BB11JFMt.

Hilgevoord, Jan, and Jos Uffink. "The Uncertainty Principle." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016. https://plato.stanford.edu/archives/win2016/entries/qt-uncertainty/.

hiQ Labs, Inc. v. LinkedIn Corporation, No. 17–16783 (9th Cir. September 9, 2019).

International Telecommunication Union. "Statistics." *International Telecommunication Union*. Accessed June 30, 2020. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

Jones, Meg Leta. "Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication." *Communication Law and Policy* 23, no. 2 (April 3, 2018): 159–95. https://doi.org/10.1080/10811680.2018.1430418.

Kaiser, Anna Jean. "The Brazilian Group Scanning WhatsApp for Disinformation in Run-up to Elections." *The Guardian*, September 26, 2018, sec. World news. https://www.theguardian.com/world/2018/sep/26/brazil-elections-comprova-project-misiniformation-whatsapp.

Kaufmann, Katja, and Corinna Peil. "The Mobile Instant Messaging Interview (MIMI): Using WhatsApp to Enhance Self-Reporting and Explore Media Usage in Situ." *Mobile Media & Communication* 8, no. 2 (May 1, 2020): 229–46. https://doi.org/10.1177/2050157919852392.

Kemp, Simon. "Digital 2020: Global Digital Overview." DataReportal in collaboration with HootSuite and WeAreSocial, January 30, 2020. https://datareportal.com/reports/digital-2020-global-digital-overview.

Klosowski, Thorin. "What's the Deal with All These Messaging Apps?" *Lifehacker*, April 10, 2014. https://lifehacker.com/whats-the-deal-with-all-these-messaging-apps-1561543034.

Korean Ministry of Science and ICT. "2018 년 인터넷이용실태조사 결과 발표 (Announcement of the 2018 Internet Usage Survey)." *대한민국 정책포털 KOREA.KR 정책브리핑 (Korea Policy Portal KOREA.KR Policy Briefing)*, February 25, 2019. http://www.korea.kr/news/pressReleaseView.do?newsId=156318833.

Lyons, Kim. "Google Is Indexing WhatsApp Group Chat Links, Making Even Private Groups Discoverable." *The Verge*, February 21, 2020. https://www.theverge.com/2020/2/21/21147073/whatsapp-google-group-chat-join-indexing-links-search-privacy-facebook.

Mac, Ryan, and Pranav Dixit. "The Tip Line WhatsApp Launched to Combat Fake News Isn't Actually Going to Combat Fake News." *BuzzFeed News*, April 3, 2019. https://www.buzzfeednews.com/article/ryanmac/whatsapp-fake-news-tip-line-indian-election-not-helpline.

Machado, Caio, Beatriz Kira, Vidya Narayanan, Bence Kollanyi, and Philip Howard. "A Study of Misinformation in WhatsApp Groups with a Focus on the Brazilian Presidential Elections." In *Companion Proceedings of The 2019 World Wide Web Conference*, 1013–19. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3308560.3316738.

McMahon, Jordan. "Why We Should All Ditch Other Messaging Apps for Signal." *Wired*, November 5, 2017. https://www.wired.com/story/ditch-all-those-other-messaging-apps-heres-why-you-should-use-signal/.

Mehta, Ivan. "[Best of 2019] WhatsApp Finally Lets You Prevent People from Adding You to Their Shitty Groups." *The Next Web*, April 3, 2019. https://thenextweb.com/apps/2019/04/03/whatsapp-finally-lets-you-prevent-people-from-adding-you-to-their-shitty-groups/.

Mello, Patrícia Campos. "Empresários bancam campanha contra o PT pelo WhatsApp." *Folha de S.Paulo*, October 18, 2018, sec. Poder. https://www1.folha.uol.com.br/poder/2018/10/empresarios-bancam-campanha-contra-o-pt-pelo-whatsapp.shtml.

Melo, Philipe, Johnnatan Messias, Gustavo Resende, Kiran Garimella, Jussara Almeida, and Fabrício Benevenuto. "WhatsApp Monitor: A Fact-Checking System for WhatsApp." *Proceedings of the International AAAI Conference on Web and Social Media* 13 (July 6, 2019): 676–77.

"Mobile Services — Bots, the next Frontier." *The Economist*, April 9, 2016, sec. Business. https://www.economist.com/business/2016/04/09/bots-the-next-frontier.

Mont'Alverne, Camila. Letter to Aleksei Kaminski. "E-Mail Reply: Introdução e Sua Pesquisa- Aleksei Kaminski," March 18, 2020.

———. Letter to Aleksei Kaminski and Connie Moon Sehat. "Re: Oi e Conselho! - Aleksei," July 23, 2020.

Mont'Alverne, Camila, and Isabele Mitozo. "Muito Além da Mamadeira Erótica: As notícias compartilhadas nas redes de apoio a presidenciáveis em grupos de WhatsApp, nas eleições brasileiras de 2018 (Beyond the Erotic Baby Bottle: News shared through WhatsApp by presidential candidates' support networks during the Brazilian elections in 2018)." In *Compolítica8* 25 (2019). http://ctpol.unb.br/compolitica2019/GT4/gt4_Montalverne_Mitozo.pdf.

Moreno, Andrés, Philip Garrison, and Karthik Bhat. "WhatsApp for Monitoring and Response during Critical Events: Aggie in the Ghana 2016 Election." In *Proceedings of the 14th ISCRAM Conference*, 11. Albi, France, 2017.

Mpofu, Shepherd, and Shanade Bianca Barnabas. "Citizen Journalism and Moral Panics: A Consideration of Ethics in the 2015 South African Xenophobic Attacks." *African Journalism Studies* 37, no. 4 (November 2016): 115–36. https://doi.org/10.1080/23743670.2016.1256053.

Napoli, Philip M. "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers." *Telecommunications Policy* (Special Issue on the Governance of Social Media) 39, no. 9 (October 1, 2015): 751–60. https://doi.org/10.1016/j.telpol.2014.12.003.

Narayanan, Vidya. Letter to Connie Moon Sehat. "Re: [Request for Input] Research Practices for Closed Messaging Groups," July 28, 2020.

Narayanan, Vidya, Bence Kollanyi, Ruchi Hajela, Ankita Barthwal, Nahema Marchal, and Philip N. Howard. "News and Information over Facebook and WhatsApp during the Indian Election Campaign." Computational Propaganda Project Data Memo. Oxford Internet Institute, May 13, 2019. https://comprop.oii.ox.ac.uk/research/india-election-memo/ (now at https://comprop.oii.ox.ac.uk/research/posts/news-and-information-over-facebook-and-whatsapp-during-the-indian-election-campaign/).

Newman, Nic. "Reuters Institute Digital News Report 2020." Reuters Institute for the Study of Journalism, 2020. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.

Newman, Nic, Richard Fletcher, Antonis Kalogeropoulos, David A. L. Levy, and Rasmus Kleis Nielsen. "Reuters Institute Digital News Report 2018." Reuters Institute for the Study of Journalism, 2018.

Ohm, Paul. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review* 57, no. 6 (August 2010). https://www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/.

Omanga, Duncan. "WhatsApp as 'Digital Publics': The Nakuru Analysts and the Evolution of Participation in County Governance in Kenya." *Journal of Eastern African Studies* 13, no. 1 (January 2, 2019): 175–91. https://doi.org/10.1080/17531055.2018.1548211.

Oyebode, Oluwabunmi O., and Adeyemi Adegoju. "Appraisal Resources in Select WhatsApp Political Broadcast Messages in the 2015 Presidential Election Campaign in Nigeria." *Journal of Pan African Studies* 10, no. 10 (November 1, 2017): 29.

Pamment, James. "EU Code of Practice on Disinformation: Briefing Note for the New European Commission." Partnership for Countering Influence Operations. Policy Perspectives Series. Carnegie Endowment for International Peace, March 2020. https://carnegieendowment.org/2020/03/03/eu-code-of-practice-on-disinformation-briefing-note-for-new-european-commission-pub-81187.

Pasquetto, Irene V., Eaman Jahani, Alla Baranovsky, and Matthew A Baum. "Understanding Misinformation on Mobile Instant Messengers (MIMs) in Developing Countries." *Harvard Shorenstein Center on Media, Politics and Public Policy*, May 2020. https://shorensteincenter.org/wp-content/uploads/2020/06/Misinfo-on-MIMs-Shorenstein-Center-May-2020.pdf

Paul, Emmanuel. "Everything You Need to Know about Nigeria's Social Media Bill and What You Can Do about It." *Techpoint Africa* (blog), November 28, 2019. https://techpoint.africa/2019/11/28/nigerias-social-media-bill/.

Pedro Burgos. "What 100,000 WhatsApp Messages Reveal about Misinformation in Brazil." *First Draft* (blog), June 27, 2019. https://firstdraftnews.org:443/latest/what-100000-whatsapp-messages-reveal-about-misinformation-in-brazil/.

Perrigo, Billy. "The Inside Story of How Signal Became the Private Messaging App for an Age of Fear and Distrust (Updated)." *Time*, September 28, 2020. https://time.com/5893114/signal-app-privacy/.

Proto. "PROTO | Checkpoint 'FAQs About Project Checkpoint.'" Accessed July 8, 2020. https://www.checkpoint.pro.to/.

Rennó, Raquel. "WhatsApp: The Widespread Use of WhatsApp in Political Campaigning in the Global South (Updated Version)." *Our Data Our Selves (Tactical Tech)* (blog), November 13, 2018. https://ourdataourselves.tacticaltech.org/posts/whatsapp/.

Resende, Gustavo, Philipe Melo, Julio C. S. Reis, Marisa Vasconcelos, Jussara M. Almeida, and Fabrício Benevenuto. "Analyzing Textual (Mis)Information Shared in WhatsApp Groups." In *Proceedings of the 10th ACM Conference on Web Science*, 225–34. WebSci '19. Boston, Massachusetts, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3292522.3326029.

Resende, Gustavo, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. "(Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures." In *The World Wide Web Conference on* -

WWW '19, 818–28. San Francisco, CA, USA: ACM Press, 2019. https://doi.org/10.1145/3308558.3313688.

Samuelson, Pamela. "Enriching Discourse on Public Domains." *Duke Law Journal* 55, no. 4 (February 1, 2006): 783–834.

Sandvig v. Barr, No. 16–1368 (D.D.C. March 27, 2020).

Saurwein, Florian, and Charlotte Spencer-Smith. "Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe." *Digital Journalism* 8, no. 6 (July 2, 2020): 820–41. https://doi.org/10.1080/21670811.2020.1765401.

Schaefer, Bruno, Tiago Barbosa, Sara Epitácio, and Roberta Resende. "Qual o Impacto Do WhatsApp Em Eleições? Uma Revisão Sistemática (2010-2019)." *Revista Debates* 13, no. 3 (December 16, 2019): 58–88. https://doi.org/10.22456/1982-5269.96255.

Sehat, Connie Moon. "Closed Messaging Spaces: Thinking About Best Practices and Ethics in a Global Environment." *Medium* (blog), October 4, 2018. https://misinfocon.com/closed-messaging-spaces-thinking-about-best-practices-and-ethics-in-a-global-environment-10fc82191743.

Signal Support. "Group Chats." Accessed February 5, 2021. https://support.signal.org/hc/en-us/articles/360007319331-Group-chats.

Solano Gallego, Esther. "How Bolsonaro Came to Pass." *Rosa Luxemburg Stiftung NYC* (blog), November 27, 2018. http://www.rosalux-nyc.org/how-bolsonaro-came-to-pass/.

Sumartono, Suwandi Sumartias. "Utilization of Whatsapp As a Political Communication Channel Politicians Padang City." Conference Paper. In *Advances in Social Science, Education and Humanities Research* 136:428–31. Atlantis Press, 2017. https://doi.org/10.2991/icosop-17.2018.67.

Telegram. "Supergroups 5000: Public Groups, Pinned Posts," March 13, 2016. https://telegram.org/blog/supergroups5k.

Tandoc, Edson C, Darren Lim, and Rich Ling. "Diffusion of Disinformation: How Social Media Users Respond to Fake News and Why." *Journalism* 21, no. 3 (March 2020): 381–98. https://doi.org/10.1177/1464884919868325.

tlgrm.eu. "Telegram Channels List: Discover Interesting Channels for Your Telegram." Accessed Jan. 22, 2021. https://tlgrm.eu/channels

Telegram. "Telegram FAQ." Accessed January 11, 2021. https://telegram.org/faq#q-how-secure-is-telegram.

----------. "Telegram Privacy Policy." Telegram. Accessed February 19, 2021. https://telegram.org/privacy.

"The Gambia." Freedom on the Net 2020. Freedom House, 2020. https://freedomhouse.org/country/gambia/freedom-net/2020.

Timberg, Craig, and Drew Harwell. "Pro-Trump Forums Erupt with Violent Threats Ahead of Wednesday's Rally against the 2020 Election." *Washington Post*, January 5, 2021. https://www.washingtonpost.com/technology/2021/01/05/parler-telegram-violence-dc-protests/.

United Nations General Assembly. "Universal Declaration of Human Rights (UDHR)," December 10, 1948. https://www.un.org/en/universal-declaration-human-rights/index.html.

Uttam, Kumar. "For PM Modi's 2019 Campaign, BJP Readies Its WhatsApp Plan." *Hindustan Times*, September 29, 2018. https://www.hindustantimes.com/india-news/bjp-plans-a-whatsapp-campaign-for-2019-lok-sabha-election/story-lHQBYbxwXHaChc7Akk6hcI.html.

U.S. Department of Health and Human Services. Federal Policy for the Protection of Human Subjects ('Common Rule') (2018). https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

Vaas, Lisa. "Google Stops Indexing WhatsApp Chats; Other Search Engines Still at It." *Naked Security* (blog), February 25, 2020. https://nakedsecurity.sophos.com/2020/02/25/google-stops-indexing-whatsapp-chats-other-search-engines-still-at-it/.

Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media + Society* 6, no. 1 (January 1, 2020). https://doi.org/10.1177/2056305120903408.

Wardle, Claire, Angela Pimenta, Guilherme Conter, Nic Dias, and Pedro Burgos. "Comprova: An Evaluation of the Impact of a Collaborative Journalism Project on Brazilian Journalists and Audiences." *First Draft*, June 27, 2019. https://firstdraftnews.org:443/latest/comprova-an-evaluation-of-the-impact-of-a-collaborative-journalism-project-on-brazilian-journalists-and-audiences/.

We Are Social. "2018 Digital Yearbook." Internet, 17:11:28 UTC. https://www.slideshare.net/wearesocial/2018-digital-yearbook-86862930.

WhatsApp. "Giving More Time for Our Recent Update." *WhatsApp.Com*, January 15, 2021. https://blog.whatsapp.com/giving-more-time-for-our-recent-update/?lang=en.

———. "Keeping WhatsApp Personal and Private." *WhatsApp Blog*, April 7, 2020. https://blog.whatsapp.com/Keeping-WhatsApp-Personal-and-Private.

———. "More Changes to Forwarding." *WhatsApp Blog*, January 21, 2019. https://blog.whatsapp.com/more-changes-to-forwarding.

———. "Two Billion Users — Connecting the World Privately." *WhatsApp Blog*, February 12, 2020. https://blog.whatsapp.com/two-billion-users-connecting-the-world-privately.

———. "WhatsApp Encryption Overview: Technical White Paper [Version 3]," October 22, 2020. https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf.

———. "WhatsApp FAQ — How to Use Broadcast Lists." Accessed July 13, 2020. https://faq.whatsapp.com/iphone/chats/how-to-use-broadcast-lists.

———. "WhatsApp Legal Info." Accessed July 8, 2020. https://www.whatsapp.com/legal/#key-updates.

Wiesenberg, Markus, and Ralph Tench. "Deep Strategic Mediatization: Organizational Leaders' Knowledge and Usage of Social Bots in an Era of Disinformation." *International Journal of Information Management* 51 (November 26, 2019). https://doi.org/10.1016/j.ijinfomgt.2019.102042.

Zommer, Laura. Letter to Connie Moon Sehat. "[Chequeado Statistics on WhatsApp and Misinformation]," December 23, 2020. https://chequeado.com/latamcoronavirus/.

Additional Resources

First Draft. *How Journalists Can Monitor TikTok to Inform Their Reporting | Webinar with Laura Garcia*, 2020. https://www.youtube.com/watch?v=n1WVwPQcwRY&list=PLon8am2uBRCBUZI5ftm2jQ63tlYrtY4_I&index=10.

———. *Reporting on Coronavirus - Using WhatsApp to Find Covid-19 Stories | Webinar*, 2020. https://www.youtube.com/watch?v=mUfUqI8a7bs&list=PLon8am2uBRCBUZI5ftm2jQ63tlYrtY4_I&index=2.

Al Khaja, Khalid A. J., Alwaleed K. AlKhaja, and Reginald P. Sequeira. "Drug Information, Misinformation, and Disinformation on Social Media: A Content Analysis Study." *Journal of Public Health Policy* 39, no. 3 (August 1, 2018): 343–57. https://doi.org/10.1057/s41271-018-0131-2.

Melo, Philipe de Freitas, Carolina Coimbra Vieira, Kiran Garimella, Pedro O. S. Vaz de Melo, and Fabrício Benevenuto. "Can WhatsApp Counter Misinformation by Limiting Message Forwarding?" In *Proceedings of the International Conference on Complex Networks and Their Applications*. Lisbon, Portugal, 2019. http://arxiv.org/abs/1909.08740.

Reis, Julio C. S., Philipe Melo, Kiran Garimella, Jussara M. Almeida, Dean Eckles, and Fabrício Benevenuto. "A Dataset of Fact-Checked Images Shared on WhatsApp During the Brazilian and Indian Elections." *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 26, 2020): 903–8.

Resende, Gustavo, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. "(Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures." In *The World Wide Web Conference*, 818–828. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3308558.3313688.

Tsarwe, Stanley. "How Apps on Mobile Phones Are Changing Zimbabwe's Talk Radio." *The Conversation* (blog), August 2, 2020. http://theconversation.com/how-apps-on-mobile-phones-are-changing-zimbabwes-talk-radio-143611.

Acknowledgements

Many thanks to Avery Davis-Roberts, Kiran Garimella, Jamie Hitchen, Marie Bohner, and Victoria Kwan for suggestions and feedback, as well as the participants of the October 2019 and 2020 MisinfoCon Elections workshops. The authors want to especially thank participants from the Montreal AI Ethics Institute, who kindly reviewed an earlier draft of the paper and also provided detailed suggestions. Finally, we greatly appreciate the feedback from investigators cited.

This paper would not have been possible without the kind support of the MisinfoCon Election 2019 and 2020 workshop sponsors, the Mozilla Foundation and Craig Newmark Philanthropies. Tattle would also like to express its appreciation to the AI Ethics Grant for enabling work on WhatsApp.

As mentioned at the outset, past workshops in 2018 and 2019 laid the foundation for this paper's inquiry. Calls for suggestions of related case studies, as well as for researchers who may be interested in collaborating, were issued to the same workshop participants in addition to the Association of Internet Researchers and Responsible Data Forum group email lists in the first half of 2020. We greatly appreciate the clarifications regarding research practices that we received from many of the investigators whose studies were cited under the four models.